CREATE
Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

http://create.canterbury.ac.uk

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g. Piper, K. (2014) Interpretation of clinical imaging examinations by radiographers: a programme of research. Ph.D. thesis, Canterbury Christ Church University.

Contact: create.library@canterbury.ac.uk

Canterbury Christ Church University

Interpretation of clinical imaging examinations by radiographers: a programme of research

by

Keith Piper

Canterbury Christ Church University

Thesis submitted for the
Degree of Doctor of Philosophy

2014

Interpretation of clinical imaging examinations by radiographers: a programme of research

Abstract

Background

Studies which have investigated the interpretation of plain skeletal examinations by radiographers have demonstrated encouraging findings, however, the studies have not extended beyond this area of practice and radiographers' diagnostic performance for other more complex investigations has not been established. Comparisons of performance between groups of healthcare practitioners to date, has also been limited.

Aim

This research programme aimed to investigate the interpretation of clinical imaging examinations by radiographers, and other healthcare practitioners, in the provision of initial interpretations and/or definitive reports of plain imaging (skeletal and chest) and cross-sectional (magnetic resonance imaging [MRI] – lumbar/thoracic spine, knees and internal auditory meati [IAM]) investigations.

Methods

The eight studies utilised a variety of methodological approaches and included quasi-experimental and observational studies. One quasi-experimental study compared the performance of radiographers, nurses and junior doctors in initial image interpretation and another similar study included a training intervention; both utilised alternate free-response receiver operating characteristic curve (AFROC) methodology. Three of the observational studies investigated the ability of radiographers to provide definitive reports on a wide range of clinical examinations, including chest and MRI investigations, in a controlled environment. One large multi-centre observational study investigated the performance of radiographers, in clinical practice (A/E: skeletal examinations) during the implementation of a radiographic reporting service. The agreement between consultant radiologists' MRI reports of lumbar/thoracic spine, knee and IAM examinations was investigated in another observational study. The final study compared the reports of trained radiographers and consultant radiologists, with those of an index radiologist, when reporting on MRI examinations of the knee and lumbar spine, as part of a prospective pre-implementation agreement study.

Results

The first AFROC study demonstrated statistically significant improvements after training, for radiographers (A_1=0.55 - 0.72) and nurses (A_1=0.65 - 0.63), although the radiographers maintained a better overall performance post training ($p=0.004$) in providing an initial image interpretation of trauma radiographs of the appendicular skeleton. Radiographers also achieved statistically higher ($p<0.01$) AUC values (A_1=0.75) than nurses (A_1=0.58) and junior doctors (A_1=0.54) in the second AFROC study.

Three studies, which examined 11155 reports, were conducted under controlled conditions in an academic setting and provided evidence of radiographers' high levels of accuracy in reporting of skeletal A/E (93.9%); skeletal non A/E (92.5%); chest (89.0%); MRI lumbar/thoracic spine (87.2%), knees (86.3%) and IAM (98.4%) examinations.

In the multi-centre clinical study, the mean accuracy, sensitivity and specificity rates of the radiographers reports (n=7179) of plain examinations of the skeletal system in the trauma setting was found to be 99%, 98% and 99%, respectively.

The considerable range of values for agreement, between consultant radiologists reports of MRI examinations of the thoracic/lumbar spine (k=0 – 0.8), knee (k=0.3 – 0.8) and IAM (k=1.0) was similar to other studies and resulted in a reasonable estimation of the performance, in the UK, of an average non specialist consultant radiologist in MRI reporting.

In the final study, radiographers reported in clinical practice conditions, on a prospective random sample of knee and lumbar spine MRI examinations, to a level of agreement comparable with non-musculoskeletal consultant radiologists (Mean difference in observer agreement <1%, $p=0.86$). Less than 10% of observers' reports (radiographers and consultant radiologists) were found to be sufficiently discordant to be clinically important.

Conclusion

The outcomes of this research programme demonstrate that radiographers can provide initial interpretations of radiographic examinations of the appendicular skeleton, in the trauma setting, to a higher level of accuracy than A/E practitioners. The findings also provide evidence that selected radiographers with appropriate education and training can provide definitive reports on plain clinical examinations (A/E and non A/E referral sources) of the skeletal system and the chest; and MRI examinations of the knee, lumbar/thoracic spine and IAM to a level of performance comparable to the average non specialist consultant radiologist. Wider implementation of radiographer reporting is therefore indicated and future multi-centre research, including economic evaluations, to further inform practice at a national level, is recommended.

Acknowledgements

This research would not have been possible without the help of the following and my grateful thanks are extended to them all:

- The radiographers, radiologists and other healthcare practitioners that participated in the studies;
- The PhD supervisory team for their helpful comments and suggestions;
- Professor Audrey Paterson, for her unwavering inspirational guidance and support;
- My wife, Andrea; and children, Matthew, Sam and Lauren for their love and understanding.

Contents Page

Section 3

List of figures

List of appendices:

The studies listed on the following pages were originally included as ANNEX 1 to the thesis submitted for examination.

Study 1 Piper, K. J. and Paterson, A. (2009)

'Initial image interpretation of appendicular skeletal radiographs: a comparison between nurses and radiographers',

Radiography, 15(1), pp. 40-48.

Study 1 Coleman, L. and Piper, K. (2009)

'Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers',

Radiography, 15(3), pp. 196-202.

Study 3 Piper, K. Paterson, A., and Ryan, C. (1999)

'The implementation of a radiographic reporting service for trauma examinations of the skeletal system in 4 NHS trusts'

NHS Executive South Thames funded research project. Canterbury Christ Church University (then College).

Study 4 Piper, K. J., Paterson, A. M. and Godfrey, R. C. (2005)

'Accuracy of radiographers' reports in the interpretation of radiographic examinations of the skeletal system: a review of 6796 cases',

Radiography, 11(1), pp. 27-34.

Study 5 Piper, K., Cox, S., Paterson, A., Thomas, A., Thomas, N., Jeyagopal, N. and Woznitza, N. (2014)

'Chest reporting by radiographers: Findings of an accredited postgraduate programme',

Radiography, 20(2), pp. 94-99.

Study 6 Piper, K. and Buscall, K. (2008)

'MRI reporting by radiographers: The construction of an objective structured examination',

Radiography, 14(2), pp. 78-89.

Study 7 Piper, K., Buscall, K. and Thomas, N. (2010)

'MRI reporting by radiographers: Findings of an accredited postgraduate programme',

Radiography, 16(2), pp. 136-142.

Study 8 Brealey, S., Piper, K., King, D., Bland, M., Caddick, J., Campbell, P., Gibbon, A., Highland, A., Jenkins, N., Petty, D. and Warren, D. (2013)

'Observer agreement in the reporting of knee and lumbar spine magnetic resonance (MR) imaging examinations: Selectively trained MR radiographers and consultant radiologists compared with an index radiologist',

European Journal of Radiology, 82(10), pp. e597-e605.

Investigation of radiographers' contribution to the practice of interpreting clinical imaging examinations.

The structure of this commentary follows the current University guidance for submission of PhD by Publication.

Section 1

1.0 Introduction

The programme of research discussed in this commentary investigated radiographers' contributions in the field of clinical image interpretation. The work provided evidence, which was novel at the various times of publication, on radiographers' ability to report on a wide range of diagnostic imaging investigations including cross sectional and complex examinations. Studies 1 and 2 investigated the diagnostic performance of radiographers in comparison with other professional groups to provide an initial interpretation of radiographic examinations of the appendicular skeleton of patients referred from Accident and Emergency (A/E) Departments or Minor Injuries Units (MIU). Study 3 was an NHS Funded multi-centre study which investigated the implementation of a Radiographic Reporting Service (RRS) in five clinical centres in England. Two subsequent studies (Studies 4 and 5) analysed the diagnostic performance of radiographers to provide definitive reports on radiographic examinations of the appendicular and axial skeleton; and the chest. In both studies the examinations included were of patients referred from A/E and non A/E sources.
The three remaining studies related to interpretation of magnetic resonance imaging (MRI) examinations. Study 6 compared reports produced by consultant radiologists for a number of different anatomical structures of the lumbar spine and knee during the construction of an objective structured examination (OSE). Study 7 analysed the results for the first groups of radiographers who completed the OSE and Study 8 examined the potential implementation of MRI radiographer reporting into practice.

The studies included as Annex 1, were the first of their kind and with no evidence of their replication to date. The report and papers have been cited 76 times to date.

Over 250 healthcare practitioners have participated in the eight studies which involved the investigation of the reports of 18971 imaging examinations.

The concept map in Appendix 1 illustrates the interrelationship between the studies and outlines how this programme of research has made a significant impact on radiographic practice and contributed towards new professional knowledge in the field.

2.0 Rationale

Many researchers have considered the diagnostic performance of healthcare professionals to interpret clinical imaging examinations, although the majority of earlier publications have referred to the accuracy of radiologists, including the work of Birkelo et al (1947); Garland (1949); Ledley and Lusted (1959); Bland et al (1969); Fineberg (1977); Metz (1986); Elmore et al (1994); Jarvik (2001); and van Rijn et al (2005), for example. Such articles, which have focussed on chest, hand, computed tomography (CT) head, magnetic resonance imaging (MRI) lumbar spine and mammogram reporting, have reported high error and disagreement rates between observers.
Robinson (1997) noted that the weakest link, the 'Achilles Heel', in the chain of clinical imaging events was the performance of the observer. Kundel (2006, p.404) agreed and also commented that 'The history of observer performance suggests that this is a problem that is not going to go away'. As Berlin (2007, p.1176) reiterated, 'radiologists have not yet been successful in elucidating and correcting the factors involved in causing radiologic errors. Their efforts to do so will undoubtedly continue for many years to come'. This challenge is no longer the sole domain of the radiologist however, as other healthcare professionals, particularly radiographers, are now increasingly responsible for interpreting the image.

Since the introduction of radiographer reporting in the United Kingdom (UK) in the 1990s, the evidence base has expanded and now includes research which has examined the contribution of radiographers. One of the most rigorous explorations, which related to the accuracy of plain radiograph reporting in clinical practice, provided compelling evidence to support these developments (Brealey et al, 2005a). Research which has investigated radiographer reporting of other clinical imaging examinations is lacking and therefore the diagnostic performance and/or diagnostic impact or outcome has not been established for others areas.

In order to investigate whether radiographers were able to contribute to a reporting service for a range of clinical examinations and situations, this programme of research focussed on three areas in which there was a lack of evidence about the wider contribution. The projects related to the diagnostic performance of radiographers and other healthcare professionals when providing a preliminary interpretation in the trauma situation; the diagnostic performance of radiographers in the provision of definitive reports of plain skeletal or chest examinations (including patients referred from non-A/E sources); and the diagnostic performance/diagnostic outcome of radiographer reporting of internal auditory meati (IAM), lumbar/thoracic spine and knee magnetic resonance imaging (MRI) investigations, including the implementation into clinical practice. A more detailed understanding in these areas of practice provided evidence that helped to inform subsequent research and service delivery developments.

This reflective commentary critically analyses each study and demonstrates the diagnostic performance of radiographers in the reporting of a wide range of plain and cross sectional imaging examinations.

3.0 The Research Context

Sections 3.1 and 3.2 provide a background to the frameworks which have been developed to evaluate medical tests, including the interpretation of clinical imaging examinations. Section 3.3 is a historical perspective on reporting by radiographers.

3.1 Assessment of diagnostic imaging

The value of diagnostic imaging has been acknowledged for over a century (Rowland, 1896) and merely months after the initial discovery of x-rays, the benefits were being realised (Schuster, 1896). The importance of the role of the observer in the process was also soon appreciated and within 5 years, an American clinician commented that 'experience and a skilled eye are needed in reading them (radiographs) as much as a technique in making them' (Leonard, 1900, p. 164). A clinical case was described which involved the x-ray diagnosis of a patient with urinary calculus where the method (technique) was judged to be correct but the 'interpretation was inaccurate' (p. 166). Smith similarly recognised that the failure to make a positive diagnosis on a known case of renal calculus was 'due perhaps more

to an error in the interpretation of the skiagraph than due to the skiagraph itself' [1](1904, p.751). These are possibly the earliest reported incidences of errors of interpretation.
It was 40 years until the issue of evaluating diagnostic tests appeared again in the literature. In the early 1940s, and facing the demand for a mass radiography national screening programme to detect tuberculosis, the need for a comprehensive evaluation of the methods available to diagnose active pulmonary lesions was realised by the Veterans Administration in the USA (Birkelo et al, 1947). A Board of Roentgenology was established in 1944, to investigate the relative diagnostic efficiency of various techniques, for example: 35mm, 4 x 5 inch or 4 x 10 inch photofluorograms or 17 x 14 inch celluloid films. The data included interpretations by three experienced radiologists of 5000 radiographic examinations of the chest and the results were surprising. Garland found 'to his astonishment that not only did he differ from his colleagues in apparently simple interpretations, but that he even differed with himself in a significant percentage of the same films which he read on two separate occasions' (1949, p.309-310). The level of disagreement was 30% and 21% for inter-individual and intra-individual observations, respectively.
Several additional reports (Yerushalmy et al, 1950; Yerushalmy, 1955) provided further corroboration of these findings in chest interpretation which were also confirmed by other researchers, including Groth-Peterson et al (1952) and Cochrane and Garland (1952) in Denmark and England, for example. High levels of inter observer variation had also been noted in the interpretation of radiographic examinations of the cervical spine (Bland et al, 1965) and the hand (Bland, 1969).
In his address to the meeting of the New York Roentgen Society in 1959, Garland's summary included the following statement: 'The accuracy of many diagnostic procedures is subject to impairment from errors in technique of examination and interpretation' (1960, p.583). The address also included comments on the potential benefits of dual readings to more accurately identify active pulmonary tuberculosis and he speculated that from the public health point of view the gain would justify this additional inconvenience. Although not proposed as such, Garland had in fact referred to all the key elements of what would later come to be known as an evaluative framework for medical tests, which included; technical performance, diagnostic performance, diagnostic impact, therapeutic impact and impact on health and society (Fineberg, 1978).

[1] Skiagraph (from the Greek word for shadow) , x-ray, radiograph, radiogram were all terms used to describe a photographic image produced on a radiosensitive surface by radiation other than visible light - especially by X-rays or gamma rays (Hyperdictionary, 2014)

3.2 Evaluation of medical tests

In a 1977 study, which investigated the effect of cranial computerised tomography (CT) on diagnostic and therapeutic plans, Fineberg suggested that the evaluation of diagnostic imaging should be considered at the following four levels of efficacy: Level 1 – Technical Output; Level 2 – Diagnostic (and Prognostic) Information; Level 3 – Therapeutic Plan; and Level 4 – Patient Outcome (1977). He further added that 'a new medical technology could be evaluated along eight dimensions: (1) technical performance; (2) clinical efficacy; (3) resource costs, charges and efficiency; (4) safety; (5) acceptability to patients, physicians and other users; (6) research benefits for the future; (7) larger effects on the organization of health services; and (8) larger effects on society' (Fineberg, 1978, p.1).

The Institute of Medicine (1977) extended Fineberg's four levels to an evaluative framework of five levels and made the distinction between diagnostic performance of imaging and its impact on diagnostic thinking of clinicians. Based on earlier work (Ledley and Lusted, 1959; and Lusted, 1971), in 1972 the American College of Radiologists formed an Efficacy Studies Committee which distinguished between diagnostic efficacy, labelled E1, and defined as the 'influence of the radiographic information on the diagnostic thinking of the clinician'; the therapeutic effect (E2), the influence on clinical management; and outcome efficacy (E3), 'was the patient better off as a result of the procedure having been performed' (Loop and Lusted, 1978, p.174). The study also demonstrated that the x-ray examinations had an impact on the diagnostic thinking of the clinician in 92% of situations.

Guyatt et al (1986) advocated a stepwise schema to the clinical evaluation of diagnostic technologies, based on the work of Fineberg, in a 6 stage hierarchy as follows: 1) technologic capability; 2) range of possible uses; 3) diagnostic accuracy; 4) impact on health care providers; 5) therapeutic impact; and 6) patient outcome. In the UK, Freedman suggested that evaluation studies were analogous to either Phase II or Phase III of clinical trials noting that the majority of studies at that time had focussed on diagnostic accuracy and more emphasis was needed on those studies which evaluated the contribution to clinical management (1987). Fryback and Thornbury (1991) then extended the Loop and Lusted model (1978) and proposed a 6 tiered hierarchical model outlined in Figure 1.

Figure 1 Hierarchical model of efficacy

Level	Efficacy	Typical measure of analyses (illustrative)
1	Technical efficacy	Resolution of line pairs Modulation transfer function change Gray-scale range Amount of mottle Sharpness
2	Diagnostic accuracy efficacy	Yield of normal or abnormal diagnoses in a case series Diagnostic accuracy (percentage correct diagnoses in case series) Predictive value of positive or negative examination (in a case series) Sensitivity and specificity in a defined clinical problem setting Measures of ROC curve height or area under the curve
3	Diagnostic thinking efficacy	% of cases in a series in which image judged 'helpful' to making the diagnosis Change in differential diagnosis probability Difference in clinicians pre and post-test diagnosis probabilities
4	Therapeutic efficacy	% of times image judged 'helpful' in planning patient management % of times medical procedure avoided due to image information % of times therapy planned pre-test changed after image information was obtained % of times therapeutic choices changed after test information
5	Patient outcome efficacy	% of patients improved with test compared to those without Morbidity (or procedures) avoided after having image information Change in quality-adjusted life expectancy Expected value of test information in quality-adjusted life years (QALYs)
6	Societal efficacy	Benefit – cost analysis from societal viewpoint Cost – effectiveness from societal viewpoint

Adapted from Fryback and Thornbury (1991, p.90)

Mackenzie and Dixon (1995) argued that the terms efficacy and effectiveness are often used synonymously and referred to the definitions provided by the Health Technology Assessment (HTA) Advisory Group in the UK. At that time HTA was defined as the assessment of the costs, effectiveness and broader impact of all methods used by health professionals to promote health, prevent and treat disease, and improve rehabilitation and long term care (Department of Health, 1992). Mackenzie and Dixon (1995) described efficacy as the relationship between the technology and its effects in ideal conditions. Effectiveness is the extent to which the technology, in routine circumstances leads to a change in diagnosis, management plans and improvement in health. Efficiency is a financial concept associated with the optimal use of resources (1995). The current definition has altered little in the current Technology Evaluation Programme of the National Institute for Health Research (NIHR) which is now responsible for evaluations into the efficacy, effectiveness, costs and broader impact of healthcare interventions and states that 'By "technology" we mean any method used to promote health, prevent and treat disease and improve rehabilitation or long-term care. "Technologies" in this context are not confined to new drugs but include procedures, devices, tests, settings of care, screening programmes and any intervention used in the treatment, prevention or diagnosis of disease' (NHS, 2011). Technology evaluations are now under the remit of the HTA programme, the Efficacy and Mechanism Evaluation (EME) Programme or the Systematic Reviews Programme depending on the type/stage of proposed research.

Mackenzie and Dixon (1995) had previously proposed a five stage evaluative hierarchical framework, as applied to the assessment of the effects of MRI, which included technical performance, diagnostic performance, diagnostic impact, therapeutic impact and impact on health. This framework was further adapted by Brealey (2001) to assess the effects of image reporting, and specifically related to reporting by radiographers. Specific questions were posed at each of the 6 levels proposed; technical competence, diagnostic performance, diagnostic outcome, therapeutic outcome, patient outcome and societal level, as illustrated in Figure 2.

Each of the levels will be discussed with specific consideration for reporting by radiographers.

Figure 2 The evaluative hierarchy used to assess the effects of image interpretation as illustrated by specific questions related to reporting by radiographers

Technical competence

Do radiographers use visual search patterns comparable with that of an expert?

↓

Diagnostic Performance

Do radiographers accurately interpret radiographs compared with a reference standard?
Do radiographers consistently agree with the expert observers in clinical practice?

↓

Diagnostic Outcome

Does radiographer reporting:

a) Improve clinician's diagnostic confidence and understanding?
b) Displace the need for other professionals?
c) Displace the need for further investigations?
d) Complement the existing process?

↓

Therapeutic Outcome

Does radiographer reporting contribute to the planning and delivery of therapy?

↓

Patient Outcome

↓

Does radiographer reporting result in the improved health of the patient?

Societal Level

Is the cost (borne by society as a whole) of radiographer reporting acceptable?

Adapted from Brealey (2001).

3.2.1 Technical Competence

The first question raised in Brealey's framework was that of technical competence and specifically, 'Do radiographers use visual search patterns comparable with that of an expert?' (2001, p. 343), and therefore have the potential for reporting diagnostic images. The visual search behaviour of radiologists was initially investigated using eye movement tracking methods by Kundel and La Follette (1972) during the interpretation of chest radiographs. A number of subsequent studies have further investigated this aspect in relation to chest radiology, including the work of Nodine and Kundel (1987), Kundel et al (1991) and Samuel et al (1995). Krupinski (1996) initially explored the search patterns of radiologists during the interpretation of mammograms and a subsequent study included technologists (Nodine at al, 1996). In the UK, Carr and Mugglestone (1997) reported that radiographers, when viewing chest radiographs under experimental conditions, demonstrated comparable search strategies to radiologists and concluded that a case could be made for the radiographers' role to be extended into the area of interpretation and reporting. Subsequent research by Manning et al (2003) has investigated eye tracking activity particularly related to the detection of pulmonary nodules and the effect of feedback. Experienced radiographers have been included in some of the eye tracking studies; but as part of an 'experienced' observer group and not as a group distinct from radiologists. Manning et al (2006a and 2006b) investigated the nature of expert performance in the detection and localisation of significant pulmonary nodules in plain radiographic examinations of the chest and found that after training in chest interpretation (6 month module including 30 hours of formal lectures and 500 practice cases) the radiographers performance was equal to that of the radiologist (Area Under Curve ~0.8). The authors emphasised however that single pathology assessments such as this should be interpreted with caution because of the likelihood that the performance may be 'task specific'. It is difficult to comment on the difference between the radiographers post training and the radiologists, as this does not appear to have been explicitly tested except for the saccadic amplitude (Mean values: radiologists, 6.7; and radiographers, 4.5) which was found to be significantly different (t-test $p=0.0005$ and ANOVA $F=14.4$, $F_{crit}=3.11$, $p=0.00047$) suggesting that the radiologists cover the visual scene in longer sweeping movements. In terms of other eye tracking parameters; mean number of fixations per film, visual coverage and mean scrutiny time per film there appears to be little difference between the two groups based on the magnitude and overlap of the error bars included. It is interesting to note that Manning et al acknowledged that there was no convincing evidence to suggest that the teaching of fixation patterns might be beneficial.

Brealey (2001) similarly suggested that if selectively trained radiographers can effectively report radiographs in clinical practice to a level of accuracy similar to that of radiologists, the search patterns and ability to detect abnormalities are likely to be similar. Robinson (2001) agreed and questioned if technical and diagnostic competence in image interpretation can be separated. Arguably, although detection rates have been proven to be similar (for a range of diagnostic examinations), search pattern analyses have mainly been related to pulmonary nodule detection. Whilst analysis of visual search patterns might reflect the potential for radiographer reporting, arguably this stage has been omitted by many researchers and superseded by diagnostic performance and/or diagnostic outcome studies in a number of anatomical areas/types of diagnostic examination.

It should also be noted that evaluation frameworks are most likely not a linear but a cyclic and repetitive process (Lijmer, 2009).

3.2.2 Diagnostic Performance

The diagnostic performance level proposed by Brealey (2001) corresponds to Level Two of the Fineberg (1977) model (Diagnostic [and Prognostic] Information) and Level Two of the Fryback and Thornbury (1991) hierarchy (Diagnostic Accuracy Efficacy). It also aligns with the Diagnostic Performance level proposed by Mackenzie and Dixon and (1995) but differentiates between the measurement of radiographers' performance under controlled conditions, such as in an Objective Structured Examination (OSE) and measurement in clinical practice. Since the early conception of reporting courses for radiographers in the 1990s, levels of accuracy have been measured in controlled conditions and compared to a 'reference standard' (Prime at al, 1999). Brealey argues that the use of a double/triple-blind consultant radiologist report should produce valid results to assess radiographers' abilities. Previously, 95% had been proposed as an acceptable level of accuracy, and adopted by some universities, and whilst this may be considered appropriate for some aspects of reporting it is probably unrealistic for others. Robinson et al (1999a) reported that the variation between experts was considerable, noting concordance rates between all three readers (experienced consultant radiologists) of 51%, 61% and 74% for abdominal, chest and skeletal radiographs, respectively. He also noted that the disagreement rates between pairs of observers, when only major disagreements were considered, were similar for all areas (10-12%) and estimated the average incidence of errors per observer to be between 3% and 6%; and recommended that these figures be taken into account when designing assessment techniques for image reporting. It should be noted however, that Robinson's study, which examined the variation in plain film examinations of patients referred from the accident and emergency department, also found that chest and abdomen radiographs led to a wider range of descriptive and interpretative observations and in turn, a wider variation for these anatomical areas. In addition, he commented on the lack of robust methodologies for the assessment of cognitive tasks in medicine, for example, the interpretation of radiographs. Robinson had also previously commented (1997), when referring specifically to the reporting performance of observers from a non-medical background, that further research was required for more complex areas of reporting.

Concordance between consultant radiologists is generally higher when reporting plain film examinations as compared to more complex/cross-sectional investigations. Agreement when interpreting MRI investigations is known to be lower: knee 68-96% (Bryan et al, 2001); (Piper and Buscall, 2008, Study 6, Annex 1); (Piper et al, 2010, Study 7, Annex1); and lumbar spine 78-99% (Piper and Buscall, 2008, Study 6, Annex 1); (Piper et al, 2010, Study 7, Annex1) and MRI head 84% (McCarron et al, 2006). More recently, Briggs et al (2008) reported disagreement rates of 13% (major discrepancy) and 32% (minor discrepancy) when comparing the opinions of general radiologists and neuroradiologists when interpreting CT or MRI neurological investigations. There were similar rates of minor and major discrepancies for both CT and MRI examinations. Taking into account the range of percentage agreement rates demonstrated for different modalities and / or anatomical areas the OSE pass mark for agreement has, at some universities, been set at levels other than 95%. A number of the studies in this commentary (Studies 4, 5 and 7; Annex1) have investigated the diagnostic performance of radiographers in a controlled environment as part of an OSE and specifically in relation to the reporting of: plain film skeletal and chest; and MRI lumbar/thoracic spine, knee and internal auditory meati (IAM) examinations. Other studies documented in the literature, that have considered the developing performance of radiographers in clinical practice as part of education and training programmes or experimental studies, include for example, evidence related to mammography (Pauli et al, 1996; and Wivell et al, 2003); skeletal reporting (Carter and Manning, 1999); gastric screening (Yatake et al, 2009); and neurological MRI examinations of the head and cervical spine (Piper, 2009).

The measurement of radiographers' reporting performance of plain film skeletal examinations in clinical practice has been extensively examined by Brealey at al (2005a).

His meta-analysis of 12 studies, which included 29868 examinations, demonstrated that radiographers report plain radiographs in clinical practice at 92.6% sensitivity and 97.7% specificity. The largest study, which was also considered to be the highest quality (Brealey, 2005a, p.236) of the 12 studies included, was conducted at this Higher Education Institution (HEI) and is included as Study 3 of this commentary (Annex 1). This multi-centre NHS funded study investigated the implementation of the then new reporting role into clinical practice and the findings were encouraging. The project demonstrated that, for skeletal examinations (n=7168) of patients referred from the A/E Department, radiographers could report to a high level of diagnostic performance: accuracy; 99.1%; sensitivity; 97.6%; and specificity; 99.3% (Piper et al, 1999). In the final analyses, which included 10275 reports, the accuracy was confirmed to be 99% (Piper and Paterson, 2000; Paterson and Piper, 2000; Piper et al, 2000). Brealey et al (2003) had also previously found no significant differences between the area under the receiver operating characteristic (ROC) curves for radiographers and consultant radiologists when reporting A&E or GP plain radiographs.

Study 4 in this commentary (Annex 1) which analysed 6796 examinations reported by radiographers as part of an OSE, demonstrated only negligible differences between appendicular and axial examinations (Piper et al, 2005). In a more recent analysis of 27800 skeletal examinations, although no significant difference was demonstrated ($p=0.41$) between the sensitivity rates, the appendicular specificity and agreement rates were both significantly higher ($p<0.001$) than the corresponding rates for the axial skeleton cases (Piper, 2012).

The studies included in this commentary have made a significant contribution to the evidence base on radiographer reporting of plain film (skeletal and chest) and MRI thoracic/lumbar spine, knee and IAM examinations. Other examples of studies related to radiographer reporting and conducted in clinical practice are: Murphy et al (2002) who compared radiographer and radiologist reports for barium enema examinations; Bates et al (1994) who audited the role of the sonographer in non-obstetric ultrasound; and Ripsweden et al (2009) who investigated the potential contribution radiographers may offer in the interpretation of cardiac CT investigations.

3.2.3 Diagnostic, Therapeutic and Patient Outcome

With reference to diagnostic outcome, one of the specific questions raised by Brealey (2001) was: Does radiographer reporting improve clinician's diagnostic confidence and understanding? This question is largely unanswered with regard to radiographer reporting with a few exceptions, for example, Woznitza and Piper (2011, 2012a) and Woznitza et al (2013; 2014a; 2014b) and the following project, which is in progress and registered on the UK Clinical Research Network Study Portfolio: 'Establishing the diagnostic accuracy of radiographer chest x-ray reports and their influence on clinicians' clinical reasoning: A comparison with consultant radiologists' (Woznitza, 2012b). Although Brealey et al (2005b) did not seek to answer this precise question, they did find that the introduction of radiographer reporting did not have an adverse effect on A&E radiograph reporting accuracy, patient management, or outcome; and at no additional cost. It has also been demonstrated that the introduction of radiographers in a reporting role can improve the availability of reports for A&E (Piper et al, 1999, Study 3, Annex 1) and General Practitioner (GP) examinations (Brealey and Scuffham, 2005). Brealey and Scuffham also suggested that 'this is important ……for referring clinicians' decision-making and ultimately patients' health' (p. 542). More recently, Hardy et al (2013a; 2013b) have investigated the impact of immediate reporting by radiographers within the emergency department (ED) and found that the immediate reporting service resulted in a reduction in ED interpretive errors and prevented errors that would require patient recall. Immediate reporting however did not eliminate ED interpretative errors or change the number of patients discharged, referred to hospital clinics or admitted overall, and concluded that further work was needed to explore the reasons for this.

When Robinson et al (1999b) analysed 11000 skeletal examinations reported by radiographers, for the incidence of patient re-attendance, they found no detectable adverse consequences. The authors concluded however that the radiographers training was limited to the reporting of A/E examinations and that 'the extent to which the initial success of this study can be extended to unselected plain radiographic examinations and to more complex imaging studies is uncertain (Robinson at al, 1999b ,p. 550).

3.2.4 Societal level impact

The sixth and 'most global level of efficacy' proposed by Fryback and Thornbury (1991, p.89), and as outlined in Figure 1, is that of societal efficacy. At Level 6 it is proposed that the efficacy of diagnostic imaging is questioned; and in particular is the overall cost borne by society acceptable in relation to the benefits? Mackenzie and Dixon (1995) commented that the questions of whose costs to assess, including those borne by society and government, and how to evaluate the effects, remain unanswered. In relation to image interpretation, Brealey suggests a cost-benefit analysis could be undertaken to examine 'social efficiency' (2001, p.345). Although to date, no detailed cost-benefit analyses, to appraise the use of resources in the wider context, have been conducted, recent research demonstrates how radiologists and radiographers, working together, have produced significant clinical service improvements, in terms of reporting activity and report waiting times (Woznitza, 2014b).

The research programme presented in this commentary considered the contribution radiographers can make to a reporting service for these more challenging areas of interpretation. The specific aims and objectives of the research programme are included in Section 4.

3.3 Reporting by radiographers: a historical perspective

Reporting on radiographic examinations by radiographers is not a new phenomenon (Larkin, 1983) and the role of radiographers in reporting has been debated since the beginning of the last century (Arthur and Muir, 1909). Prior to the NHS Community Care Act (1990) and the subsequent relaxation of professional boundaries, the reporting of diagnostic imaging examinations was previously, almost exclusively, the domain of radiologists. The continuing rise in demand for radiological services in the early 1990s, led to the idea of using radiographers to alleviate radiologists' workloads by developing them to report on some categories of films, and whilst this had been raised previously (Swinburne, 1971) the notion was proposed again, and perhaps, most notably by Saxton (1992).

The relentless growth of workload within departments of clinical radiology and, possibly, the influence of Saxton's comments (1992) led to a number of developments in radiographer reporting. These were the introduction of an accredited postgraduate education programme for radiographers to enable them to report musculo-skeletal radiographs (Canterbury Christ

Church [then College], 1994) – work described as 'ground breaking' by McConnell, Eyres and Nightingale (2005, p.11); the evaluation of an in-service training programme at a District General Hospital to assess its impact on fracture reporting by radiographers (Loughran, 1994); and the establishment of a collaborative research project to investigate the feasibility of plain film reporting by radiographers following a structured postgraduate training programme (Wilson, 1995).

All three developments and their consequences received much attention from the radiological and radiographic communities, and each demonstrated that radiographers were capable of reporting radiographs of the musculo-skeletal system to a very high standard.

The development of the postgraduate programme at Canterbury and the initial findings were presented widely both nationally and internationally (Davies et al, 1994; Field-Boden and Piper, 1995 and 1996; McMillan al, 1995; Piper, 1995, 1996, 1997; Piper and Paterson, 1997a, 1997b; Piper et al, 1998, 1999 [Study 3, Annex 1 of this commentary]; Piper and Paterson, 2000; Piper at al, 2000; Paterson & Piper, 1999 and 2000). In particular, Piper and Paterson (1997b) examined the sensitivity and specificity rates achieved by a small group of trained radiographers who collectively reported 6592 musculo-skeletal examinations in a pre-implementation trial. The scores for both measures were greater than 97%.

Robinson (1996) assessed the accuracy of a group of trained radiographers and found that there was no significant difference when compared to a group of radiologists. In relation to fracture detection, as part of an in-house programme, Loughran (1994) found that radiographers, who had received training, improved over a six month period, with error rates reducing from 8% to 5.3%; and sensitivity and specificity rates increasing to 95.9% and 96.8%, respectively. These rates were compared to those of the radiologists who achieved rates of 96.8% (sensitivity) and 99.6% (specificity). Loughran also found (1996) that the standards of reporting by the radiographers in the study were maintained and improved over an eight month period. In a review of 5566 A/E radiographs the non-concurrence rate, when compared with a radiologist's report, reduced from 4.6% at the start of the period to 3.2% at the end.

Piper et al (1999), as outlined in Section 3.2.2, went on to investigate the implementation of a Radiographic Reporting Service (RRS) for trauma examinations of the skeletal system, in an NHS Funded multi-centre study in England. Seven thousand one hundred and seventy-nine

(7179) reports produced by radiographers were verified by an experienced radiologist and equivocal examinations were reported by a second radiologist who was blind to the radiographer's and first radiologist's reports. The mean accuracy, sensitivity and specificity rates were 99.1%, 97.6%, and 99.3%, respectively.

The practice of radiographer reporting is now recognised (RCR and SCoR, 2012), and common place; in a major survey 53% of 108 hospitals in the UK confirmed that musculo-skeletal reporting was being carried out by radiographers (SCoR, 2008).

Radiographers now also interpret more complex images of other anatomical areas and/or of diagnostic examinations from cross sectional imaging modalities, although there is less evidence available of radiographers' diagnostic performance when reporting these more complex areas (RCR and SCoR, 2012).

Radiographers have been responsible for providing initial interpretations on imaging examinations since the 1980s and most recently the Society and College of Radiographers (2013) recommended that the previously used 'red-dot system' be phased out and replaced by an initial interpretation now referred to as a ' Preliminary Clinical Evaluation'.

The studies included in this commentary have investigated radiographers' performance in the context of preliminary and definitive reporting.

4.0 Aim and objectives:

Aim

The research programme included in this commentary aimed to investigate the diagnostic performance of radiographers to report, or provide an initial interpretation, on plain and cross sectional imaging examinations on patients referred from a wide range of referral sources. The specific objectives were:

- to examine the effect of a short training programme on nurses and radiographers, exploring differences between their performance before and after training (Study 1, Annex 1);
- to assess how accurately and confidently casualty officers, nurse practitioners and radiographers, practicing within the emergency department (ED), recognise and

describe radiographic trauma within a test bank of 20 appendicular radiographs (Study 2, Annex 1);

- to evaluate the implementation of a Radiographic Reporting Service (RRS) in four NHS Trusts in the United Kingdom (Study 3, Annex 1);
- to analyse the results achieved in Objective Structured Examinations (OSEs) by a number of radiographers (n=28) who successfully completed a postgraduate qualification in clinical reporting of the appendicular and axial skeleton; and to test for any significant differences, in terms of sensitivity, specificity and/or accuracy between cases of patients referred from the accident and emergency (A/E) department when compared to other referral sources (Study 4, Annex 1);
- to analyse the objective structured examination (OSE) results of the first six cohorts of radiographers (n =40) who successfully completed an accredited postgraduate programme in clinical reporting of adult chest radiographs (Study 5, Annex 1);
- to measure the agreement between three independent radiologists reports during the construction of a bank of general magnetic resonance imaging (MRI) investigations. The bank was subsequently to be used to assess radiographers' ability to accurately report at the end of an accredited programme; Postgraduate Certificate (PgC) Clinical Reporting (MRI-General Investigations) – (Study 6, Annex 1);
- to analyse the objective structured examination (OSE) results of the first three cohorts of radiographers (n=39) who completed an accredited postgraduate certificate (PgC) programme in reporting of general magnetic resonance imaging (MRI) investigations and (for a representative sample) to compare the agreement rates with those demonstrated for a small group of consultant radiologists (Study 7, Annex 1);
- to assess agreement between trained radiographers and consultant radiologists compared with an index radiologist when reporting on magnetic resonance imaging (MRI) examinations of the knee and lumbar spine and to examine the subsequent effect of discordant reports on patient management and outcome (Study 8, Annex 1).

5.0 Sequence, process and coherence of studies included

A research concept map, which is included as Appendix 1, details the sequence, process and relationship between the studies included in this research programme. The series of published studies demonstrate my ongoing professional development as an autonomous researcher having led on six of the eight studies included.

Studies 1 and 2 focussed on the performance of radiographers, and other healthcare practitioners, to provide an initial image interpretation (now termed PCE) in the context of the trauma setting. Study 3 was a large multicentre funded study which examined the implementation into practice of a reporting service provided by radiographers. Studies 4 and 5 investigated the diagnostic accuracy of radiographers at the end of an accredited postgraduate programme on clinical reporting of skeletal or chest examinations. Study 6 investigated the agreement between consultant radiologists during the construction of an OSE which was developed to assess radiographers at the end of an MRI reporting programme. In Study 7, and at the end of the MRI reporting programme, the diagnostic accuracy of the radiographers was analysed and a representative sub sample compared for agreement with the radiologists reports produced as part of Study 6. Study 8 was a pre-implementation study in which radiographers with postgraduate education and training reported in clinical practice conditions on specific MRI examinations of the knee and lumbar spine to a level of agreement comparable with non-musculoskeletal consultant radiologists.

My relative contributions in Studies 1-8 (Annex 1) are included in Appendix 2.

Studies 3, 4, 5, 6, 7 and 8 remain prodigious, as research studies in these areas were, and still are, almost non-existent. Studies 1 and 2 are similarly unique in terms of the different professional groups included and/or the research methodology utilised.

Section 2

6.0 Critical Appraisal of included studies

The studies included investigated the diagnostic accuracy of imaging reports provided by healthcare practitioners, all studies were of an observational nature and adopted cross-sectional study or survey designs. The specific checklists used to appraise the studies are included in relation to each section as appropriate.

6.1 Study 1 (Annex 1)

Piper, K. J. and Paterson, A. (2009)
'Initial image interpretation of appendicular skeletal radiographs: a comparison between nurses and radiographers',
Radiography, 15(1), pp. 40-48.

The critical appraisal of this study was informed by the Standards for Reporting of Diagnostic Accuracy (STARD) checklist which was first published in 2003 (Bossuyt et al, 2003) and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool which was developed soon after (Whiting, 2003). The intention was that STARD checklist be used as a prospective tool to guide and develop a well-designed study, whereas the QUADAS tool be used to critique and review studies previously completed (Cook et al, 2007). Some aspects of Study 1 cannot easily be assessed using either checklist however, and a number of specific questions and issues related to the assessment of the methodological quality of diagnostic accuracy and performance studies on radiographer reporting, have been raised (Brealey and Glenny, 1999; Brealey and Scally, 2001; Brealey et al, 2002a, 2002b). These concerns have led to the formulation of more applicable criteria (Brealey, 2004; Brealey et al, 2005). Potential biases, due to selection (film and observer), the application of the reference standard, the measurement of results and independence of interpretation have been adapted, following previous work (Kelly et al, 1997) for specific use in plain film reading performance studies (Brealey and Scally, 2001).

It should also be noted, that Study 1 (Annex 1) commenced in 2002 and data collection was in progress by the time many of the tools and checklists referred to above were widely disseminated. The criteria developed by Brealey are included in Appendix 3 and the completed checklist for Study 1 is included as Appendix 4. Applying the criteria developed by

Brealey, this study can be described as a 'Cohort A versus Cohort B versus reference standard' diagnostic accuracy study conducted outside of clinical practice in controlled conditions (Brealey, 2004, p.2). As can be seen from the completed checklist the majority of criteria were met successfully, including the use of a sample size calculation as estimated by Obuchowski (2000). Although the difference in the area under the curve (AUC) values between the two groups of observers in the study (radiographers and nurses) was significant pre-training ($p=0.038$), this was not the case in the post training phase ($p=0.159$). This may have been a result of under powering the study. Initial calculations were based on very limited data available at the time as no previous studies had compared radiographers and nurses. The moderate differences which were anticipated were not realised in the study where only small differences ($< 10\%$) were evident.

The utilisation of the alternate free-response receiver operating characteristic (AFROC) methodology was unique in a study of this nature. The value of receiver operating characteristic (ROC) curves in the clinical evaluation of diagnostic tests has long been recognised as the curves produced allow a comprehensive assessment of accuracy for both true-positive and false-positive fractions for a range of threshold values (Hanley and McNeil, 1982; Metz, 1986; Weinstein et al, 2005). A limitation of the ROC methodology however is the issue of lesion localisation and diagnosis of multiple abnormalities as traditional ROC analyses do not penalise these incorrect decisions and may result in an overestimation of the accuracy of the diagnostic test being investigated. Location ROC (LROC) and free-response ROC (FROC) are two methods which have been developed to address these shortcomings (Starr, 1975; Swensson, 1996). A computer-based FROC model, which scores all clinically relevant correct decisions as true-positive and all other as false-positive was developed by Chakraborty (1989) and Chakraborty and Winter (1990). The alternate FROC (AFROC) method, which can utilise simpler ROC software after the data has been appropriately rescored, was a further development of the ROC paradigm (Chakraborty, 2002). It is estimated that the use of the FROC methodology can lead to an increase in power by a factor of 1.6 when compared to the ROC method (Chakraborty, 2002).

6.2 Study 2 (Annex 1)

Coleman, L. and Piper, K. (2009)

'Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers',

***Radiography,* 15(3), pp. 196-202.**

This quasi-experimental study was assessed using the criteria developed by Brealey (2005a) and as described in 6.1 and included in Appendix 3. The completed checklist for Study 2 is included as Appendix 5. The particularly novel aspects of this study are threefold. Firstly, no previous study had compared these three groups of healthcare practitioners when providing an initial image evaluation in the emergency department setting; secondly the study utilised AFROC methodology; and thirdly no previous study had explicitly investigated any association between the performance of these practitioners and the confidence of their decision. This study was conducted soon after Study 1 and although a larger sample size would have been preferable, this was a convenience sample influenced by the local clinical setting. Post-hoc analysis however, revealed that the actual difference in AUC values, between the group of radiographers and the nurses or doctors was almost 20% and therefore represented a large difference; >15% (Obuchowski, 2000). The large inter observer variability (>10%) suggests that an image bank of 20 examinations and 10 observers in each group would adequately satisfy the criteria to achieve an 80% power of sample with a 5% possibility of a Type I error (Obuchowski, 2000, p.604). As discussed earlier, the specific methodologies used in ROC analysis have been continually developed. During the data collection phase of this study, the free- response paradigm of location specific observer performance was further enhanced with the introduction of the jack-knife free-response (JAFROC) method (Chakraborty & Berbaum, 2004). Although the sample size for Study 2 was adequate, the results may have realised increased statistical power if the JAFROC method had been utilised.

Studies 1 and 2 have both provided important evidence to the profession and the imperative to increase the extent of image interpretation undertaken by radiographers is now professional body policy and becoming accepted practice in the UK. Both studies have been cited in articles which have considered this practice in European (Smith and Reeves, 2009; Buissink, 2014); South African (Hlongwane and Pitcher, 2013; Makanjee et al, 2014; Williams, 2013); Australian (McConnell and Smith, 2007; McConnell et al, 2012 and 2013; Brown and Leschke, 2012; Yielder, 2014) and Canadian (Hilkewich, 2014) publications.

When relating this development to Australian practice, McConnell noted this has,

> ‘gained such momentum that the College of Radiographers expects all radiographers will provide descriptions of abnormality location and characterisation on radiographs. In support of this, U.K. research (Piper & Paterson, 2005; and Coleman & Piper, 2009) has also demonstrated that radiographers are the best alternative professional group to perform this function despite the drive towards the use of the emergency nurse practitioner in minor injuries presentations’ (2013, p.49).

Studies utilising ROC methodologies to compare the performance of these different professional groups had not been conducted previously. The findings, which demonstrated the mean values, 95% confidence intervals and the levels of inter-observer variation for radiographers, nurses and junior doctors, in terms of AUC, sensitivity and specificity, were therefore unique and added new knowledge to the field. Given the number of observers and cases included, these could be regarded as Phase II studies (Obuchowski, 2004) The false positive and false negative errors which occurred most frequently were also identified and discussed at a time when the growing use of written preliminary evaluations was being advocated by the College of Radiographers (2006).

It is evident that Studies 1 and 2, which have been collectively cited 29 times, have both influenced the practice of radiography and have made an international impact.

6.3 Study 3 (Annex 1)

Piper, K. Paterson, A., and Ryan, C. (1999)

'The implementation of a radiographic reporting service for trauma examinations of the skeletal system in 4 NHS trusts'

NHS Executive South Thames funded research project. Canterbury Christ Church University (then College).

The project report is included in Annex 1. A longitudinal study design was used to measure the productivity and effectiveness of radiographic reporting in four NHS Trusts and five clinical sites in England. Data were collected by direct measure, report pro-forma, semi-structured questionnaires and interviews. A series of base line measurements was made at the commencement of the project. These were the volume of reporting activity prior to implementation of a Radiographic Reporting Service (RRS) and the speed with which the reports became available. The satisfaction of the users of the reporting service prior to the implementation of an RRS was also gauged. Three measures (volume and speed of report availability; and the satisfaction of users) were repeated after the RRS had been implemented. Longitudinal data on the accuracy of the radiographers' reports in terms of sensitivity and specificity were also collected at each site. Finally, some cost information related to the introduction and provision of an RRS was gathered.

Four NHS Trusts (five clinical centres) and 10 radiographers participated in the study. Radiographers completed 10275 reports and at the time the report (Study 3, Annex 1) was published 7179 cases had been reviewed by a radiologist and were included in the initial report to assess accuracy, sensitivity and specificity. Of the initial 7179 cases, 7074 were judged to be correct resulting in an overall accuracy at that time of 98.54%. Volume and speed data were obtained from the normal workload in each Trust. Four radiology services managers provided the cost data, while 26 staff took part in the initial survey and 12 in the final survey.

Much of the data included in this report were evaluated as part of an independent systematic review and meta-analysis which aimed to determine the diagnostic performance of radiographer plain film reporting in clinical practice (Brealey, 2005a). Searches conducted by Brealey identified 952 potential studies and after initial screening for eligibility, 927 studies were excluded from the meta-analysis. Eleven studies had not been conducted in clinical practice; seven were not accuracy studies or had incomplete data to identify TN, TP, FN and

FP fractions; nine did not assess radiographers or were case studies; four were visual search studies; and 11 were duplicate or more compete data sets were available. Twelve studies were included in the meta-analysis. All studies were reviewed by Brealey and two co-authors for study eligibility and there was perfect agreement between reviewers. The criteria used to assess each study for methodological quality have been outlined in Section 6.1 and are included as Appendix 3 of this commentary.
At the time of the meta-analysis, 7148 of the 7179 RRS cases (Study 3, Annex 1) had been verified; 7074 were judged correct and accuracy was estimated to be 98.96%. Of the 12 studies included in the meta-analysis by Brealey it is notable that the data provided from the RRS study scored highly. The mean quality score of all the studies included in the meta-analysis was 45.8, and the RRS study (Study 3, Annex 1) scored 33.2 ranking it as the highest quality diagnostic performance study included; the lowest scoring studies were judged to be of the highest quality. The checklist, as completed by Brealey (2004), is included as part of Appendix 6. The main study criticisms related to the possibility of verification bias, arbiter review bias and the lack of any estimates of inter observer or intra observer variability. In an ideal world these limitations could have been remedied by ensuring that the consultant radiologists reported all cases blind to the radiographers report. This would have been particularly challenging, partly due to the limited funding and the number of reports involved (>10000 examinations), and partly because at one of the clinical sites almost no A/E reporting (<1%) was being completed by consultant radiologists; one of the prime reasons for developing the reporting role and implementing the study at the outset. Additional funding may have enabled independent arbitration and double reporting of a sample of cases to estimate inter and intra observer variability but was beyond the scope of the project agreed by the NHS.

The service benefits of radiographer reporting are now well recognised and the report, included in Annex 1, has been cited by the Society and College of Radiographers in definitive guidance (2010).

Subsequent analyses completed soon after the report was submitted (Piper, 2000; and Paterson and Piper, 2000) included the final estimates of accuracy. Of the 10275 reports verified by a consultant radiologist, 151 were judged by the first radiologist to be incorrect (FN or FP). The sample of equivocal cases was independently reported by two additional consultant radiologists to provide a consensus report for comparison with the radiographers' report and the review by the initial radiologist. Of the 151 cases, 95 were confirmed by a second consultant radiologist, blinded to any previous report, as false negative or false positive resulting in a final accuracy of the radiographers' reports of 98.8%. It was interesting to note that in 37 cases, at least one radiologist agreed with the radiographers' report and in 40 cases, two radiologists agreed but disagreed with the third.

6.4 Study 4 (Annex 1)

Piper, K. J., Paterson, A. M. and Godfrey, R. C. (2005)

'Accuracy of radiographers' reports in the interpretation of radiographic examinations of the skeletal system: a review of 6796 cases',

***Radiography,* 11(1), pp. 27-34.**

This study was independently reviewed as part of the systematic review referred to previously Brealey (2004). The quality score check list as completed for the review is included as part of Appendix 6. The mean score for all diagnostic accuracy studies evaluated as part of the systematic review was 31.8 and Study 4 included in this commentary (Annex 1) achieved the lowest numerical value and therefore scored highest (Quality Score 10.1) and was ranked highest of the studies eligible for inclusion in the meta analysis.

Main criticisms were that no attempts were made to assess inter observer variability and that the possibility of arbiter bias had not been excluded. Also, there was no attempt to assess intra observer variability. No sample size calculations were offered and whilst this is a valid criticism, post hoc estimates suggest that 1548 cases (774 appendicular; and 774 axial) would have been adequate to detect a difference in accuracy of 93% and 95% (Scally and Brealey, 2003) between the types of skeletal examinations included and or the referrer (A/E or non A/E), assuming a 5% possibility of a Type I error was deemed acceptable. As 6796 cases were included (>3000 appendicular and >3000 axial examinations) in Study 4 (Annex 1) the power of the sample was likely to exceed 95%. An additional strength of the study was the utilisation of triple blind consultant radiologists' reports which were used as a valid reference standard, as advocated by Robinson (1997) and Brealey et al (2002b).

At a time when radiographer reporting was under particular scrutiny, this study helped to re-assure those considering implementing the practice, that appropriately trained radiographers were able to report appendicular and axial examinations, of patients referred from A/E and non A/E clinicians, with equal levels of sensitivity, specificity and accuracy. The study, which has been referred to in a recent Society and College of Radiographers publication (Paterson, 2010), has also been cited widely (28 citations) either in articles published, or related to studies conducted, in the UK (Hardy et al, 2008); Mainland Europe (Jackson and Henderson, 2010; Smith and Reeves, 2009; Ween et al, 2005); North America (Blakely et al, 2008); Australia (McConnell and Smith, 2007; McConnell at al, 2012 and 2013; Woznitza, 2014c); and Africa (Williams, 2006; and Onyema, 2011). Study 4 had also been cited in an earlier publication by the professional body (SCoR, 2006).

6.5 Study 5 (Annex 1)

Piper, K., Cox, S., Paterson, A., Thomas, A., Thomas, N., Jeyagopal, N. and Woznitza, N. (2014)
'Chest reporting by radiographers: Findings of an accredited postgraduate programme', *Radiography,* 20(2), pp. 94-99.

This study was conducted in an academic setting and analysed the chest reports (n=4000) completed by six cohorts of radiographers in a final OSE. Estimates of diagnostic accuracy were 95.4%, 95.9% and 89% in terms of sensitivity, specificity and agreement, respectively.

This study was reviewed using the checklist developed by Brealey (2004) to assess diagnostic accuracy studies as part of his systematic review. The completed quality score check list is included as Appendix 7. The methodology was almost identical to Study 4 and achieved a Quality Score of 11.

Main limitations were that no attempts were made to assess inter observer variability and that the possibility of arbiter bias had not been excluded. Also, there was no consideration of intra observer variability. No sample size calculations were included, however, post hoc estimates suggest that the sample size of 4000 cases would have been sufficiently powered to detect a small difference between subgroups if that had been desirable. A wide range of disease types was included in the analysis however no comparisons were made between A/E and non A/E referrals. This was felt to be less important for chest reporting as the complexity of interpretation is relatively unchanged by referral source. This paper was unique in that it investigated the diagnostic accuracy of radiographers in chest reporting at the end of an accredited postgraduate programme. The literature reviewed was extensive and comparisons with studies which have examined the diagnostic performance of consultant radiologists suggested that errors made by and/or variance between the two groups was likely to be similar (Herman et al, 1975; Robinson et al, 1999a; Potchen et al, 2000; Cascade et al, 2001; Donald and Barnard, 2012). Comparisons with other experienced observers, consultant radiologists, would have been interesting; however, this work is in progress currently.

The findings are well described and are likely to be of value to a wide audience and perhaps over time increasingly impact practice in a similar way to the skeletal work completed previously. It is possible, for example, that the Improving outcomes: a strategy for cancer (DH, 2011) and the associated National Awareness and Early Diagnosis Initiative (Richards, 2009) may increase demand for chest radiography and the added pressure on clinical services could be such that radiographer reporting might be a useful service development.

6.6 Study 6 (Annex 1)

Piper, K. and Buscall, K. (2008)
'MRI reporting by radiographers: The construction of an objective structured examination',
Radiography, 14(2), pp. 78-89.

In this study a sample of lumbar spine, knee and internal auditory meati (IAM) MRI examinations were reported by groups of radiologists as part of the construction of an OSE to be used to assess radiographers' diagnostic accuracy. At the time of publication the UK literature related to consultant radiologists' diagnostic accuracy/performance was limited and the majority of relevant studies had been completed in the USA. Study 6 has been critiqued using the STROBE (STrengthening the Reporting of Observational studies in Epidemiology) Standard (Von Elm, 2008), which is completed and included as Appendix 8. When analysed critically using this tool, the study has some shortcomings, for example, the primary purpose, which was to measure the agreement between consultant radiologists, was not clearly stated. Politically, this was intentional, as it was important to articulate that the main aim of the project was to construct a reliable bank to measure radiographers' diagnostic accuracy. Nevertheless, the findings were interesting for all practitioners that report MRI examinations, as the majority of published studies which examined the diagnostic accuracy/performance of experienced radiologists had been completed outside of the UK, and mainly in the USA. The variation found between the consultant radiologists in this study compared reasonably with the findings of previous studies which had investigated agreement in reporting of MRI examinations of the knee (Umans et al, 1995; White et al, 1997; Sonin et al, 2002; Bryan et al, 2001; and Roos et al, 2006); and lumbar spine (Brant-Zawadzki et al, 1995; van Rijn et al, 2005, 2006; Jarvik et al, 1996, 2001; Mulconrey et al, 2006; and Pfirrmann et al, 2001).

Although the sample size was a pragmatic choice with the main objective being to return a sufficient number of agreed cases to construct an OSE of 40 cases, it is likely that the power of the initial sample (n=87) was sufficient (>80%) to detect substantial agreement; k=0.8 (Landis and Koch, 1977) assuming a prevalence of 50% (Sim and Wright, 2005). No intra observer variation data were collected and arbiter bias was not considered although all cases ultimately utilised in the OSE (Study 7, Annex 1) were also verified by another consultant radiologist.

No detailed demographic information was collected on the observers involved in the study, however all were experienced observers in MRI reporting of thoracic/lumbar spine; knee and IAM investigations, and were employed as DGH consultant radiologists in one of three centres. The findings of this study were therefore likely to be reasonably representative of the 'average reader' (Obuchowski, 1996, p.518).

6.7 Study 7 (Annex 1)

Piper, K., Buscall, K. and Thomas, N. (2010)
'MRI reporting by radiographers: Findings of an accredited postgraduate programme',
***Radiography,* 16(2), pp. 136-142.**

This study was conducted in an academic setting and analysed the OSE results for three cohorts of MRI reports completed by radiographers. Estimates of diagnostic accuracy were 99.0%, 99.0% and 89.2% in terms of sensitivity, specificity and agreement respectively.

This study was reviewed using the checklist developed by Brealey (2004) to assess diagnostic accuracy studies as part of his systematic review. The completed quality score check list is included as Appendix 9. The methodology was identical to Study 5 and achieved a Quality Score of 11.

Main limitations were that no attempts were made to assess inter observer variability and that the possibility of arbiter bias had not been excluded. Also, there was no consideration of intra observer variability. Although no sample size calculations were included, post hoc estimates suggest that the sample of approximately 460 knee and 580 thoracic/lumbar spine examinations included in Study 7 would have been sufficiently powered to detect a small difference between these anatomical areas if necessary (Scally and Brealey, 2003). Assuming an accuracy of 86.3% for knee and 87.2% for lumbar/thoracic spine examinations (Study 7, Annex 1); given values for Type I error and Type II error to be 0.05 and 0.2 respectively; and accepting a 5% difference in clinical practice between the accuracy of lumbar spine and knee reports, 408 examinations in each category would have resulted in an adequately powered equivalence study (Scally and Brealey, 2003, p.244).

The range of disease type was included in the paper however no comparisons were made between different referral sources. Similar to chest reporting, this was felt to be less important than for plain film skeletal reporting as the complexity of interpretation in MRI reporting is relatively unchanged by referral source.

This paper was novel in that it investigated the diagnostic accuracy of radiographers in MRI reporting at the end of an accredited postgraduate programme. Another unique aspect to this study was the additional analyses which compared a small sub group of representative radiographer reports with the consultant radiologist reports referred to in Study 6. Kappa rates were similar for both knee and lumbar/thoracic spine examinations; irrespective of whether the group consisted solely of consultant radiologists and/or if reporting radiographers were included. For some categories, significant differences were found between agreement rates for different groups (knee: bone bruise, $p=0.0067$; effusion, $p<0.001$; and lumbar/thoracic spine: tumour/metastases, $p=0.02$; other incidental findings, $p=0.04$). In the majority of these cases (9/11; Tables 4 and 5, Study 7, Annex 1) however, this was due to a higher agreement rate between one of the radiologists' reports and the radiographers' reports suggesting that the majority of consultant radiologists were more likely to agree with the reporting radiographers than one particular group of consultant radiologists included in the study.

This study was of particular value at a time when the Royal College of Radiologists (2010) was publishing guidance to radiologists and healthcare providers on reporting by non-radiologists suggesting that radiographers would only provide descriptive reports which were likely to be of limited value and not 'clinically relevant' (RCR, 2010, p.2). This document has since been withdrawn by the RCR and current guidance published jointly with the College of Radiographers, is more positive (RCR and SCoR, 2012) although highlights the lack of research evidence for some areas of reporting.

The findings are well described and are likely to be of value to a wider audience and perhaps over time increasingly impact practice in a similar way to the skeletal work completed previously.

6.8 Study 8 (Annex 1)

Brealey, S., Piper, K., King, D., Bland, M., Caddick, J., Campbell, P., Gibbon, A., Highland, A., Jenkins, N., Petty, D. and Warren, D. (2013)

'Observer agreement in the reporting of knee and lumbar spine magnetic resonance (MR) imaging examinations: Selectively trained MR radiographers and consultant radiologists compared with an index radiologist',

***European Journal of Radiology,* 82(10), pp. e597-e605.**

Study 8 was an agreement study and although diagnostic accuracy was not estimated in terms of sensitivity or specificity, an evaluation of diagnostic outcome and potential effects on patient management was included. This study was reviewed using the checklist developed by Brealey (2004) to assess diagnostic accuracy/performance/outcome studies as part of his systematic review. The quality score check list was amended due to the non-applicability of some sections; the completed check list is included as Appendix 10.

An initial comparison of agreement between radiographers and radiologists MRI reports, which was completed as part of Study 7, was conducted in a controlled environment at the end of an accredited postgraduate programme of study using a small sample size, it not being designed primarily as an inter observer study. The sample size calculation for Study 8 resulted in a larger sample of 326 MRI examinations being included and the inclusion of a sample size calculation is a strength of this study. Another positive aspect of this study was the use of an expert musculo-skeletal (MSK) consultant radiologist who provided the index report, although arguably a double/triple blind report would have been more robust as on occasions the index radiologist may have been incorrect and may be regarded as a limitation. Nemec et al (2008) for example, found that the sensitivity of a consensus report agreed by two experienced MSK radiologists was found to be 88% when compared to arthroscopic findings in cases of meniscal tear. Van Rijn et al (2005) investigated the level of inter observer variation for the presence of herniation or bulging intervertebral discs and found that the level of agreement between two experienced neuroradiologists was 84%. The other main limitations were that no attempts were made to assess intra observer variability and that the possibility of arbiter bias was not been excluded. Other limitations, including the small number of observers, are included in the published article.

A major strength of the study was that it was conducted in the clinical setting and under normal viewing conditions. The most significant findings were that less than 10% of observer reports (radiologist or radiographer) were sufficiently discordant with the index report to be clinically important and that there were no significant differences between the proportions of discordant reports that were issued by the radiographers compared to the radiologists (Mean difference in observer agreement $<1\%$, $p=0.86$). This was the first study of this nature that compared radiologists and radiographers MRI reporting knee and lumbar spine investigations in a diagnostic outcome study and the findings are likely to be of interest to those responsible for provision of efficient and effective MRI services.

6.9 Summary of strengths and limitations

The common strengths of the research included in this commentary are: new and often unique studies which have analysed reporting by radiographers in a number of different areas of developing professional practice; novel use of particular data analysis techniques not used extensively previously; and large sample sizes (Studies 3, 4, 6 and 8).

The major common limitations are the possibility of arbiter bias; no attempts to assess inter or intra observer variation; and the consistent inclusion of sample size calculations. Sample size calculations were not performed for some studies, as these were often based on OSE results and the number of examinations included was pre-determined by the assessment requirements of the particular accredited postgraduate programme and the cohorts of observers available at the time of publication. Post hoc calculations suggest, however, that the size of the samples included would have been adequate to demonstrate significant statistical relationships where this was important. The possibility of arbiter bias was therefore a possibility, however all OSE results included in the studies were second marked consistent with university procedures and reviewed by an external examiner (consultant radiologist).

Importantly, all cases included in the OSEs had been independently reported by three consultant radiologists, and this is a major strength of much of the research included in this commentary (Robinson, 1999; and Brealey et al, 2002b).

Section 3

7.0 Contribution towards knowledge and the need for further research

To date, the report and papers included in this research programme have been referenced 76 times in journals nationally and internationally and have contributed to the developing practice of radiography during the last 20 years. The impact of the various research publications has been considerable at a time of significant change within the profession.

The findings from this research programme have contributed towards a better understanding of the diagnostic performance and, to some extent, the diagnostic outcome, of radiographers to report on diagnostic images. The results have suggested that in addition to plain radiographs of the skeletal system, radiographers can interpret plain film images of the chest and cross sectional MRI investigations to a high level of accuracy.

The initial studies demonstrated that radiographers' diagnostic accuracy (in terms of sensitivity and specificity) was higher than other healthcare practitioners who routinely interpret radiographic examinations in clinical practice, when interpreting images of the appendicular skeleton. These studies were published at a time when the Society & College of Radiographers was re-iterating earlier guidance (CoR, 1997) which recommended that within the context of an abnormality detection scheme, the 'red-dot' be replaced by an initial interpretation (SCoR, 2006). The findings highlighted areas where practice could be improved and provided re-assurance to radiographers and departments considering this practice. Study 1 highlighted the performance of radiographers and nurses before and after a short course in initial image interpretation and highlighted education and training needs for future development. Study 2 adopted a post-test only design and provided additional data on how the confidence of the practitioner related to their diagnostic accuracy. It was only for the group of radiographers where any association was significant. Studies1 and 2 were also unique at the time, as the methodology used was the AFROC paradigm which resulted in greater statistical power than the traditional ROC design.

Recent research (Hardy et al, 2013a, 2013b) demonstrated that immediate reporting (including by reporting radiographers) significantly reduced ED interpretive errors and prevented errors that would require patient recall. However, immediate reporting did not eliminate ED interpretative errors or change the number of patients discharged, referred to hospital clinics or admitted overall, suggesting that further work is indicated in this area.

Study 3 was a significant report when radiographer reporting was in its infancy and qualified reporting radiographers were being utilised in a formal reporting role in the UK for the first time since the 1920s. It was vital therefore that the practice was scrutinised in detail during implementation and this study was the largest diagnostic performance study conducted in clinical practice at the time, or to date. It provided compelling evidence, not only that radiographers could accurately report trauma examinations of the skeletal system, but also that in the small number of cases where the radiographers' report was discordant, there was a similar possibility that consultant radiologists would also disagree. The report also provided evidence of the other benefits of the radiographer reporting service in that the volume of reports completed and the speed of report availability improved significantly in the majority of hospital trusts involved.

Whilst the recent document by the RCR and SCoR (2012) notes that a multidisciplinary team approach has been demonstrated to be effective in a number of areas including MSK reporting, the evidence of radiographer reporting in other areas of plain film and cross-sectional imaging is limited. Study 4, which was a diagnostic accuracy study which compared OSE results (appendicular and axial skeleton) collected in a controlled environment, goes some way to address that deficit. The results of the first three cohorts of radiographers who completed the postgraduate qualification at this institution resulted in a large sample of cases (n=6796) which were subjected to detailed analysis, including multi-level modelling. This form of data analysis had not previously been used in studies of this nature, which have investigated the diagnostic accuracy of radiographer reports. The extensive analysis concluded that any differences between the appendicular and axial skeleton scores were likely to be negligible. Both Studies 3 and 4 have been independently evaluated and scored methodologically as the highest quality diagnostic performance and diagnostic accuracy studies, respectively, available at that time (Brealey, 2004). Both have therefore provided robust findings which confirmed the accuracy of radiographer reporting in the trauma and non-trauma setting at the time and have made significant contributions to the evidence base which has led to widespread adoption of the practice within the UK.

A multi-centre diagnostic outcome study, including robust cost-benefit analysis, to investigate radiographer reporting of patients referred by GPs could be beneficial. Radiographers have accessed the chest reporting programme at this institution since 2002 and Study 5 was, and is, the first and largest, review of OSE results undertaken (n=4000). Recent literature on radiologist reporting accuracy, performance or outcome is limited in the UK and much of the literature included in the study emanated from the USA. No previous study was found which reported on the diagnostic accuracy of radiographers to report chest radiographs at the end of an accredited postgraduate programme of study. The findings suggested that the diagnostic accuracy of radiographers was likely to be similar to non-specialist consultant radiologists but a diagnostic performance study would be valuable. It will also be worthwhile to compare the diagnostic accuracy, performance and/or outcome of radiographers and consultant radiologists.

The construction of the MRI OSE included as Study 6 was primarily an inter observer study which examined the variation between consultant radiologists in MRI reporting of thoracic/lumbar spine, knee and IAM investigations. Whilst detailed Kappa analysis to investigate variation between observers is not unique, it is usually only applied to relatively small anatomical regions or pathological conditions; the other alternative being to use overall agreement for an entire report. The resultant detailed analysis in Study 6 provided useful and unique data, on agreement rates between DGH consultant radiologists in the UK. The secondary purpose was the construction of an OSE which could then be used to test radiographers' sensitivity and specificity rates in MRI reporting. The findings confirmed the pass standard for the OSE, which will also be of interest to others involved in assessments of this nature. A multi-centre study, with a larger sample size, to examine agreement rates between DGH consultant radiologists and reporting radiographers, in comparison with MSK or neuro radiologist/s providing the index report, could provide further interesting information.

Study 7 provided unique results of the first three cohorts of radiographers who completed the first accredited MRI reporting programme for radiographers. The evidence has helped to provide a growing number of NHS trusts and independent sector companies with the confidence to implement radiographer reporting of certain categories of MRI examinations; thoracic/lumbar spine and knee. The findings were encouraging and further related work in the area of head and cervical spine MRI examinations reporting by radiographers, originally reported in 2009 (Piper, 2009) is ongoing.

Study 8 was conducted in the clinical practice setting and compared the reports of two DGH non specialist consultant radiologists and two reporting radiographers, with an MSK specialist consultant radiologist who provided the index report. Two consultant orthopaedic surgeons, specialists in knee or spinal surgery, then judged any discordant reports for the clinical importance of the difference and found that there was no statistically significant difference between the radiographers and the radiologists' reports. A multi-centre diagnostic outcome study, including robust cost-benefit analysis, to investigate this further would be beneficial.

There is no doubt that the body of work included in this commentary has made a significant contribution to the body of available evidence which in turn has helped to inform the developing practice of clinical reporting by radiographers. The following extract is part of the citation (Appendix 1, p.16) given when I was awarded the Fellowship of the College of Radiographers in 2010,

'Throughout the past 10 years this (work) has focussed on the development and delivery of image interpretation programme in relation to plain film imaging, CT scanning of the head and MR imaging. In all of these areas, radiographer-led image reporting represented ground breaking achievements for the profession. Keith has also contributed significantly to the critical and vital research base that underpins the role of image interpretation for radiographers - research for which he has received national and international critical acclaim. Without the work of Keith in relation to image interpretation it is unlikely our profession would have realized the significant advances in professional practice achieved over the past 10 years.' (Evans, 2010).

8.0 Conclusion

The programme of study included in this commentary aimed to investigate the diagnostic performance of radiographers to report, or provide an initial interpretation, on plain radiograph and cross sectional examinations on patients referred from a wide range of referral sources.

The outcomes of this ground breaking programme of research demonstrate that appropriately educated and trained radiographers can report clinical imaging examinations of a complex nature to a similar degree of diagnostic performance to a consultant DGH radiologist.

Given the increasing demand for imaging services, and the growing demands for efficiency, these findings are likely to be even more important to the development of any future provision.

Further work, much of which is in progress, includes the investigation of radiographers' interpretations of neurological MRI investigations of the brain and cervical spine (Piper, 2009); a comparison of radiographers and radiologist interpretations of plain chest examinations; and the diagnostic impact of radiographers' reports on clinicians' decision making.

References

Arthur, D. and Muir, J. (1909) *A Manual of Practical X-ray Work.* London: Heinemann.

Bates, J. A., Conlon, R. M. and Irving, H. C. (1994) 'An audit of the role of the sonographer in non-obstetric ultrasound', *Clinical Radiology,* 49(9), pp. 617-620.

Berlin, L. (2007) 'Accuracy of diagnostic procedures: has it improved over the past five decades?', *AJR,* 188(5), pp. 1173-1178.

Birkelo, C. C., Chamberlain, W. E., Phelps, P. S., Schools, P. E., Zacks, D. and Yerushalmy, J. (1947) 'Tuberculosis case finding: A comparison of the effectiveness of various roentgenographic and photofluorographic methods', *Journal of the American Medical Association,* 133(6), pp. 359-366.

Blakeley, C., Hogg, P. and Heywood, J. (2008) 'Effectiveness of UK radiographer image reading', *Radiologic Technology,* 79(3), pp. 221-226.

Bland, J. H., Van Buskirk, F. W., Tampas, J. P., Brown, E. and Clayton, R. (1965) 'A study of roentgenologic criteria for rheumatoid arthritis of the cervical spine', *American Journal of Roentgenology,* 95(4), pp. 949-954.

Bland, J. H., Soule, A. B., Van Buskirk, F. W., Brown, E. and Clayton, R. V. (1969) 'A study of inter- and intra-observer error in reading plain roentgenograms of the hands. "To err is human"', *The American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine,* 105(4), pp. 853-859.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., Lijmer, J. G., Moher, D., Rennie, D. and de Vet, H.,C.W. (2003) 'Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative', *Annals of Internal Medicine,* 138(1), pp. 40.

Brant-Zawadzki, M. N., Jensen, M. C., Obuchowski, N., Ross, J. S. and Modic, M. T. (1995) 'Interobserver and intraobserver variability in interpretation of lumbar disc abnormalities: a comparison of two nomenclatures', *Spine,* 20(11), pp. 1257-1263.

Brealey, S. and Glenny, A. (1999) 'A framework for radiographers planning to undertake a systematic review', *Radiography,* 5(3), pp. 131-146.

Brealey, S. (2001) 'Measuring the effects of image interpretation: an evaluative framework', *Clinical Radiology,* 56(5), pp. 341-347.

Brealey, S. and Scally, A. J. (2001) 'Bias in plain film reading performance studies', *BJR,* 74(880), pp. 307.

Brealey, S., Scally, A. J. and Thomas, N. B. (2002a) 'Presence of bias in radiographer plain film reading performance studies', *Radiography,* 8(4), pp. 203-210.

Brealey, S., Scally, A. J. and Thomas, N. B. (2002b) 'Review article: methodological standards in radiographer plain film reading performance studies', *BJR,* 75(890), pp. 107.

Brealey, S., K., Brealey, S., King, D. G., Crowe, M. T. I., Crawshaw, I., Ford, L., Warnock, N. G., Mannion, R. A. J. and Ethell, S. (2003) 'Accident and emergency and general practitioner plain radiograph reporting by radiographers and radiologists: a quasi- randomized controlled trial', BJR*,* 76(901), pp. 57.

Brealey, S. (2004) *An Evaluation of Radiographer Plain Radiograph Reporting*. PhD Thesis. University of York, York.

Brealey, S., Scally, A., Hahn, S., Thomas, N., Godfrey, C. and Coomarasamy, A. (2005a) 'Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis', *Clinical Radiology,* 60(2), pp. 232-241.

Brealey, S., King, D. G., Hahn, S., Godfrey, C., Crowe, M. T. I., Bloor, K., Crane, S. and Longsworth, D. (2005b) 'The costs and effects of introducing selectively trained radiographers to an A&E reporting service: a retrospective controlled before and after study', *BJR,* 78(930), pp. 499.

Brealey, S. D. and Scuffham, P. A. (2005) 'The effect of introducing radiographer reporting on the availability of reports for Accident and Emergency and General Practitioner examinations: a time- series analysis', *BJR,* 78(930), pp. 538.

Brealey, S., P., Brealey, S., Piper, K., King, D., Bland, M., Caddick, J., Campbell, P., Gibbon, A., Highland, A., Jenkins, N., Petty, D. and Warren, D. (2013) 'Observer agreement in the reporting of knee and lumbar spine magnetic resonance (MR) imaging examinations: Selectively trained MR radiographers and consultant radiologists compared with an index radiologist', *European Journal of Radiology,* 82(10), pp. e597-e605.

Briggs, G., Flynn, P., Worthington, M., Rennie, I. and McKinstry, C. (2008) 'The role of specialist neuroradiology second opinion reporting: is there added value?', *Clinical Radiology,* 63(7), pp. 791-795.

Brown, N. and Leschke, P. (2012) 'Evaluating the true clinical utility of the red dot system in radiograph interpretation', *Journal of Medical Imaging and Radiation Oncology,* 56(5), pp. 510-513.

Bryan, S., Weatherburn, G., Bungay, H., Hatrick, H., Salas, C., Parry, C., Field, S. and Heatley, F. (2001) 'The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint', *Health Technology Assessment,* 5(27), pp. 1-95.

Buissink, C., T., Buissink, C., Thompson, J. D., Voet, M., Sanderud, A., Kamping, L. V., Savary, L., Mughal, M., Rocha, C. S., Hart, G. E., Parreiral, R., Martin, G. and Hogg, P. (2014) 'The influence of experience and training in a group of novice observers: A jackknife alternative free- response receiver operating characteristic analysis', *Radiography, 20(4), p.300-305.*

Canterbury Christ Church (then College) University (1994) '*Submission for approval of Canterbury Christ Church College of a Proposal for a Postgraduate Certificate Radiography (Clinical Reporting'. Canterbury: CCCC, p. 1-90.*

Carr, D. and Mugglestone, M. (1997) 'Visual search strategies of radiographers: evidence for role extension', *European Congress of Radiology,* Vienna, 7-12 March. Eur Radiol (Suppl.1 Vol 9) pp. S120.

Carter, S. and Manning, D. (1999) 'Performance monitoring during postgraduate radiography training in reporting—a case study', *Radiography,* 5(2), pp. 71-78.

Chakraborty, D. P. (1989) 'Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data', *Medical Physics,* 16(4), pp. 561-568.

Chakraborty, D. P. and Winter, L. H. (1990) 'Free-response methodology: alternate analysis and a new observer-performance experiment', *Radiology,* 174(3 Pt 1), pp. 873-881.

Chakraborty, D. (2002) 'Statistical power in observer-performance studies: comparison of the receiver operating characteristic and free-response methods in tasks involving localization', *Academic Radiology,* 9(2), pp. 147-156.

Chakraborty, D. P. and Berbaum, K. S. (2004) 'Observer studies involving detection and localization: modeling, analysis, and validation', *Medical Physics,* 31(8), pp. 2313-2330.

Cascade, P. N., Kazerooni, E. A., Gross, B. H., Quint, L. E., Silver, T. M., Bowerman, R. A., Pernicano, P. G. and Gebremariam, A. (2001) 'Evaluation of competence in the interpretation of chest radiographs', *Academic Radiology,* 8(4), pp. 315-321.

Cochrane, A. and Garland, L. (1952) 'Observer error in the interpretation of chest films: an international investigation', *The Lancet,* 260(6733), pp. 505-509.

Coleman, L. and Piper, K. (2009) 'Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers', *Radiography,* 15(3), pp. 196-202.

Cook, C., Cleland, J., Huijbregts, P., Cook, C., Cleland, J. and Huijbregts, P. (2007) 'Creation and Critique of Studies of Diagnostic Accuracy: Use of the STARD and QUADAS Methodological Quality Assessment Tools', *The Journal of manual & manipulative therapy,* 15(2), pp. 93.

Davies, S., Piper, K., McKay, L. and Paterson, A (1994) 'The Development of a Post Graduate Diploma in Clinical Reporting (Radiography)', *Autumn Conference. Moving Boundaries*, Canterbury*:* College of Radiographers, pp. 6-7

Department of Health (1990) National Health Service and Community Care Act 1990. Available at: http://www.legislation.gov.uk/ukpga/1990/19/contents (Accessed 9 Oct 2014)

Donald, J. J. and Barnard, S. A. (2012) 'Common patterns in 558 diagnostic radiology errors', *Journal of Medical Imaging and Radiation Oncology,* 56(2), pp. 173-178.

Elmore, J. G., Wells, C. K., Lee, C. H., Howard, D. H. and Feinstein, A. R. (1994) 'Variability in radiologists' interpretations of mammograms', *New England Journal of Medicine,* 331(22), pp. 1493-1499.

Evans, R. (2010) Society awards three Fellowships at UKRC. Synergy News (July). SCoR

Field-Boden, Q. and Piper, K. (1996) 'Reporting for radiographers', *Synergy (March 1996),* pp. 32-33.

Field-Boden, Q. C. and Piper, K. (1996) 'Do radiologists really need to report on everything?', *The Radiographer: The Official Journal of the Australian Institute of Radiography,*43 (3), pp.111-3.

Fineberg, H. V., Bauman, R. and Sosman, M. (1977) 'Computerized cranial tomography: effect on diagnostic and therapeutic plans', *JAMA,* 238(3), pp. 224-227.

Fineberg, H. V. (1978) 'Evaluation of computed tomography: achievement and challenge', *AJR,* 131(1), pp. 1-4.

Freedman, L. S. (1987) 'Evaluating and comparing imaging techniques: a review and classification of study designs', *BJR,* 60(719), pp. 1071-1081.

Fryback, D. G. and Thornbury, J. R. (1991) 'The efficacy of diagnostic imaging', *Medical decision making : an international journal of the Society for Medical Decision Making,* 11(2), pp. 88-94.

Garland, L. H. (1949) 'On the Scientific Evaluation of Diagnostic Procedures: Presidential Address Thirty-fourth Annual Meeting of the Radiological Society of North America 1', *Radiology,* 52(3), pp. 309-328.

Garland, L. H. (1960) 'The problem of observer error', *Bulletin of the New York Academy of Medicine,* 36pp. 570-584.

Groth-Petersen, E., Lovgreen, A. and Thillemann, J. (1952) 'On the reliability of the reading of photofluorograms and value of dual reading', *Acta tuberculosea Scandinavica,* 26(1-2), pp. 13-37.

Guyatt, G. H., Tugwell, P. X., Feeny, D. H., Haynes, R. B. and Drummond, M. (1986) 'A framework for clinical evaluation of diagnostic technologies', *CMAJ : Canadian Medical Association journal,* 134(6), pp. 587-594.

Hanley, J. A. and McNeil, B. J. (1982) 'The meaning and use of the area under a receiver operating characteristic (ROC) curve', *Radiology,* 143(1), pp. 29-36.

Hardy, M., Snaith, B. and Smith, T. (2008) 'Radiographer reporting of trauma images: United Kingdom experience and the implications for evolving international practice', *Journal of Medical Radiation Sciences,* 55(1), pp. 16-19.

Hardy, M., Hutton, J. and Snaith, B. (2013a) 'Is a radiographer led immediate reporting service for emergency department referrals a cost effective initiative?', *Radiography,* 19(1), pp. 23-27.

Hardy, M., S., Hardy, M., Snaith, B. and Scally, A. (2013b) 'The impact of immediate reporting on interpretive discrepancies and patient referral pathways within the emergency department: a randomised controlled trial', *BJR,* 86(1021), pp. 20120112.

Herman, P. G., Gerson, D. E., Hessel, S. J., Mayer, B. S., Watnick, M., Blesser, B. and Ozonoff, D. (1975) 'Disagreements in chest roentgen interpretation.', *Chest,* 68(3), pp. 278-282.

Hilkewich, M. W. (2014) 'Written Observations as a Part of Computed Tomography Angiography Post Processing by Medical Radiation Technologists: A Pilot Project', *Journal of Medical Imaging and Radiation Sciences,* 45(1), pp. 31-36. e1.

Hlongwane, S. T. and Pitcher, R. D. (2013) 'Accuracy of after-hour'red dot'trauma radiograph triage by radiographers in a South African regional hospital', *SAMJ: South African Medical Journal,* 103(9), pp. 638-640.

Jackson, M. E. and Henderson, J. E. (2010) 'Following trauma, should adult wrist radiographic examinations be two or three projections?', *Emergency Radiology,* 17(2), pp. 87-93.

Jarvik, J. G. (2001) 'The research framework', *AJR,* 176(4), pp. 873-878.

Jarvik, J. G., Haynor, D. R., Koepsell, T. D., Bronstein, A., Ashley, D. and Deyo, R. A. (1996) 'Interreader reliability for a new classification of lumbar disk disease', *Academic Radiology,* 3(7), pp. 537-544.

Jarvik, J. J., Hollingworth, W., Heagerty, P., Haynor, D. R. and Deyo, R. A. (2001) 'The longitudinal assessment of imaging and disability of the back (LAIDBack) study: baseline data', *Spine,* 26(10), pp. 1158-1166.

Kelly, S., Berry, E., Roderick, P., Harris, K. M., Cullingworth, J., Gathercole, L., Hutton, J. and Smith, M. A. (1997) 'The identification of bias in studies of the diagnostic performance of imaging modalities', *BJR,* 70(838), pp. 1028-1035.

Krupinski, E. A. (1996) 'Visual scanning patterns of radiologists searching mammograms', *Academic Radiology,* 3(2), pp. 137-144.

Kundel, H. L. (2006) 'History of research in medical image perception', *Journal of the American College of Radiology,* 3(6), pp. 402-408.

Kundel, H. L. and La Follette Jr, Paul S. (1972) 'Visual Search Patterns and Experience with Radiological Images 1', *Radiology,* 103(3), pp. 523-528.

Kundel, H. L., Nodine, C. F. and Toto, L. (1991) 'Searching for lung nodules: the guidance of visual scanning', *Investigative Radiology,* 26(9), pp. 777-781.

Landis, J. R. and Koch, G. G. (1977) 'The measurement of observer agreement for categorical data', *Biometrics,* pp. 159-174.

Larkin, G. (1983) *Occupational monopoly and modern medicine.* Tavistock Publications London.

Ledley, R. S. and Lusted, L. B. (1959) 'Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason', *Science (New York, N.Y.),* 130(3366), pp. 9-21.

Leonard, C. L., (1900) 'II. The Technique of the Positive and Negative Diagnosis of Ureteral and Renal Calculi by the Aid of the Rontgen Rays', *Annals of Surgery,* 31(2), pp. 163-179.

Lijmer, J. G., Leeflang, M. and Bossuyt, P. M. (2009) 'Proposals for a phased evaluation of medical tests', *Medical decision making : an international journal of the Society for Medical Decision Making,* 29(5), pp. E13-21.

Loop, J. W. and Lusted, L. E. (1978) 'American College of Radiology Diagnostic Efficacy Studies', *AJR,* 131(1), pp. 173-179.

Loughran, C. F. (1994) 'Reporting of fracture radiographs by radiographers: the impact of a training programme', *BJR,* 67(802), pp. 945.

Loughran, C. (1996) 'Radiographer reporting of accident and emergency radiographs: a review of 5000 cases', *BJR,* Suppl to Vol 69, pp. 128 UKRC

Lusted, L. B. (1971) 'Decision-making studies in patient management', *The New England Journal of Medicine,* 284(8), pp. 416-424.

Mackenzie, R. and Dixon, A. (1995) 'Measuring the effects of imaging: an evaluative framework', *Clinical Radiology,* 50(8), pp. 513-518.

Makanjee, C. R., Bergh, A. and Hoffmann, W. A. (2014) "So You Are Running Between"—A Qualitative Study of Nurses' Involvement With Diagnostic Imaging in South Africa', *Journal of Radiology Nursing,* 33(3), pp. 105-115.

Manning, D., Barker-Mill, S., Donovan, T. and Crawford, T. (2006b) 'Time- dependent observer errors in pulmonary nodule detection', *BJR,* 79(940), pp. 342.

Manning, D., Ethell, S. C. and Crawford, T. (2003) *Eye-tracking AFROC study of the influence of experience and training on chest X-ray interpretation.* International Society for Optics and Photonics. San Diego, CA

Manning, D., Ethell, S., Donovan, T. and Crawford, T. (2006a) 'How do radiologists do it? The influence of experience and training on searching for chest nodules', *Radiography,* 12(2), pp. 134-142.

McCarron, M. O., Sands, C. and McCarron, P. (2006) 'Quality assurance of neuroradiology in a District General Hospital', *QJM : monthly journal of the Association of Physicians,* 99(3), pp. 171-175.

McConnell, J., Eyres, R. and Nightingale, J. (2008) *Interpreting trauma radiographs.* John Wiley & Sons.

McConnell, J. and Smith, T. (2007) 'Submission to the National Health and Hospitals Reform Commission: redesigning the medical imaging workforce in Australia', *Commonwealth of Australia,*

McConnell, J., Devaney, C., Gordon, M., Goodwin, M., Strahan, R. and Baird, M. (2012) 'The impact of a pilot education programme on Queensland radiographer abnormality description of adult appendicular musculo-skeletal trauma', *Radiography,* 18(3), pp. 184-190.

McConnell, J., Devaney, C. and Gordon, M. (2013) 'Queensland radiographer clinical descriptions of adult appendicular musculo-skeletal trauma following a condensed education programme', *Radiography,* 19(1), pp. 48-55.

McMillan, P., Paterson, A. and Piper, K. (1995) 'Radiographers' abnormality detection skills in radiography of the chest and abdomen', *Rontgen Centenary Congress,* Birmingham: British Institute of Radiology, pp. 21.

Meek, S., Kendall, J., Porter, J. and Freij, R. (1998) 'Can accident and emergency nurse practitioners interpret radiographs? A multicentre study', *Journal of Accident & Emergency Medicine,* 15(2), pp. 105-107.

Metz, C. E. (1986) 'ROC methodology in radiologic imaging', *Investigative Radiology,* 21(9), pp. 720-733.

Mulconrey, D. S., Knight, R. Q., Bramble, J. D., Paknikar, S. and Harty, P. A. (2006) 'Interobserver reliability in the interpretation of diagnostic lumbar MRI and nuclear imaging', *The Spine Journal,* 6(2), pp. 177-184.

Murphy, M., L., Murphy, M., Loughran, C. F., Birchenough, H., Savage, J. and Sutcliffe, C. (2002) 'A comparison of radiographer and radiologist reports on radiographer conducted barium enemas', *Radiography,* 8(4), pp. 215-221.

National Health Service (2011) National Institute for Health Research Efficacy and mechanism evaluation (EME) programme

Available at: http://www.nets.nihr.ac.uk/which-programme/technology (Accessed 3/4/2104)Available at: http://www.nets.nihr.ac.uk/programmes/eme (Accessed 5/9/2014)

Department of Health (2011) Improving Outcomes: a Strategy for Cancer
Available at: https://www.gov.uk/government/publications/the-national-cancer-strategy
(Accessed 9 Oct 2014)

Nemec, S. F., Marlovits, S., Trattnig, S., Matzek, W., Mayerhoefer, M. E. and Krestan, C. R. (2008) 'High-resolution magnetic resonance imaging and conventional magnetic resonance imaging on a standard field-strength magnetic resonance system compared to arthroscopy in patients with suspected meniscal tears', *Academic Radiology,* 15(7), pp. 928-933.

Nodine, C. F. and Kundel, H. L. (1987) 'Using eye movements to study visual search and to improve tumor detection', *Radiographics : a review publication of the Radiological Society of North America, Inc,* 7(6), pp. 1241-1250.

Nodine, C. F., Kundel, H. L., Lauver, S. C. and Toto, L. C. (1996) 'Nature of expertise in searching mammograms for breast masses', *Academic Radiology,* 3(12), pp. 1000-1006.

Obuchowski, N. A. and Zepp, R. C. (1996) 'Simple steps for improving multiple-reader studies in radiology', *AJR,* 166(3), pp. 517-521.

Obuchowski, N. A. and Obuchowski, N. A. (2000) 'Sample size tables for receiver operating characteristic studies', *AJR,* 175(3), pp. 603.

Obuchowski, N. A. (2004) 'How many observers are needed in clinical studies of medical imaging?', *AJR,* 182(4), pp. 867-869.

Onyema, L., (2011) 'A case for CT head and plain film reporting role for radiographers in some major UK trauma Centres and their Counterparts in the developing world', *Nigerian Journal of Medical Imaging and Radiation Therapy Vol,* 1(2), pp. 11.

Paterson, A. and Piper, K. (1999) 'The Implementation of a Radiographic Reporting Service for Skeletal Trauma', *European Congress of Radiology,* Vienna, 7-12 March. Eur Radiol (Suppl.1 Vol 9) pp. S120.

Paterson, A. and Piper, K. (2000) *'The implementation of a radiographic reporting service for skeletal trauma', Clinical Reporting by Radiographers Meeting,* Radiology Section, The Royal Society of Medicine.

Paterson, A. (2010) 'Medical image interpretation: Interprofessional teams or parallel universes', *Imaging Onc,* pp. 8-13.

Pauli, R., H., Pauli, R., Hammond, S., Cooke, J. and Ansell, J. (1996) 'Radiographers as film readers in screening mammography: An assessment of competence under test and screening conditions', *BJR,* 69(817), pp. 10-14.

Pfirrmann, C. W., Metzdorf, A., Zanetti, M., Hodler, J. and Boos, N. (2001) 'Magnetic resonance classification of lumbar intervertebral disc degeneration', *Spine,* 26(17), pp. 1873-1878.

Piper, K. (1995) 'Clinical Reporting in Radiography', *75 Years and Onwards Anniversary Conference.* University of Bath: 7-9 April. The Society and College of Radiographers, pp. 16.

Piper, K. (1996) '*Accuracy of radiographers' reports in examinations of the skeletal system*'. *Radiology UK,* 20-22 May. Birmingham: *BJR (*Suppl. to Vol 69), p. 282-283.

Piper, K. (1997) 'Reporting by Radiographers: Findings of an Accredited Postgraduate Skeletal Reporting Programme in the UK', *Radiological Society of North America Scientific Assembly,* 30 Nov-5 Dec, Chicago: Radiology (Suppl. To Vol 205), pp. 360.

Piper, K. and Paterson, A. (1997a) 'The Accuracy of Radiographers' reports in Accident and Emergency Examinations of the Skeletal System', *European Congress of Radiology,* 2-7 March, Vienna: Eur Radiol (Suppl. To Vol 7), pp. S 178-179.

Piper, K. and Paterson, A. (1997b) 'The Accuracy of Radiographers' Reports in Examinations of the Skeletal System'*, Radiology UK*, 19-21 May, Birmingham: Br J Radiol (Suppl. To Vol 70), p. 123.

Piper, K., Paterson, A. and Ryan, C. (1998) 'The Implementation of a Radiographic Reporting Service for Trauma Examinations of the Skeletal System', *Radiology UK,* 1-3 June, Birmingham: Br J Radiol (Suppl. To Vol 70), p. 123.

Piper, K. Paterson, A. and Ryan, C. (1999) *'The implementation of a radiographic reporting service for trauma examinations of the skeletal system in 4 NHS trusts'. NHS Executive South Thames funded research project.* Canterbury Christ Church University (then College).

Piper, K. and Paterson, A. (2000) 'The Implementation of a Radiographic Reporting Service', *ISRTT/AIR Radiography Conference*, 18-22 Feb, Sydney: ISRRT/AIR, pp.95.

Piper, K., Paterson, A. and Ryan, C. (2000) 'Implementation of a Radiographic Reporting Service for Skeletal Trauma Examinations: Final Analysis of Accuracy'. *IOS Congress*, 22-24 May, Birmingham: Br J Radiol (Suppl. To Vol 73), p. 91.

Piper, K. J., Paterson, A. M. and Godfrey, R. C. (2005) 'Accuracy of radiographers' reports in the interpretation of radiographic examinations of the skeletal system: a review of 6796 cases', *Radiography,* 11(1), pp. 27-34.

Piper, K. and Buscall, K. (2008) 'MRI reporting by radiographers: The construction of an objective structured examination', *Radiography,* 14(2), pp. 78-89.

Piper, K. J. and Paterson, A. (2009) 'Initial image interpretation of appendicular skeletal radiographs: a comparison between nurses and radiographers', *Radiography,* 15(1), pp. 40-48.

Piper, K. (2009) 'Reporting by radiographers: Neurological magnetic resonance imaging examinations of the head and cervical spine'. *UKRC,* 8-10 June, Manchester: Conference Proceeding, BIR, p.106.

Piper, K., Buscall, K. and Thomas, N. (2010) 'MRI reporting by radiographers: Findings of an accredited postgraduate programme', *Radiography,* 16(2), pp. 136-142.

Piper, K. (2012) *'Skeletal reporting by radiographers: a review of 27800 cases'.* UKRC, 25-27 June, Manchester: Congress Guide, p. 133.

Piper, K., Cox, S., Paterson, A., Thomas, A., Thomas, N., Jeyagopal, N. and Woznitza, N. (2014) 'Chest reporting by radiographers: Findings of an accredited postgraduate programme', *Radiography,* 20(2), pp. 94-99.

Potchen, E. J., Cooper, T. G., Sierra, A. E., Aben, G. R., Potchen, M. J., Potter, M. G. and Siebert, J. E. (2000) 'Measuring Performance in Chest Radiography ', *Radiology,* 217(2), pp. 456-459.

Prime, N., Paterson, A. and Henderson, P. (1999) 'The development of a curriculum—a case study of six centres providing courses in radiographic reporting', *Radiography,* 5(2), pp. 63-70.

Richards, M. (2009) 'The national awareness and early diagnosis initiative in England: assembling the evidence', *British Journal of Cancer,* 101, pp. S1-S4.

Ripsweden, J., Mir-Akbari, H., Brolin, E. B., Brismar, T., Nilsson, T., Rasmussen, E., Rück, A., Svensson, A., Werner, C. and Winter, R. (2009) 'Is training essential for interpreting cardiac computed tomography?', *Acta radiologica,* 50(2), pp. 194-200.

Robinson, P. J. (1996) 'Short communication: plain film reporting by radiographers--a feasibility study', *BJR,* 69(828), pp. 1171.

Robinson, P. J., Culpan, G. and Wiggins, M. (1999b) 'Interpretation of selected accident and emergency radiographic examinations by radiographers: a review of 11000 cases', *BJR,* 72(858), pp. 546.

Robinson, P. J. (1997) 'Radiology's Achilles' heel: error and variation in the interpretation of the Rontgen image', *BJR,* 70(839), pp. 1085-1098.

Robinson, P. J., Wilson, D., Coral, A., Murphy, A. and Verow, P. (1999a) 'Variation between experienced observers in the interpretation of accident and emergency radiographs', *BJR,* 72(856), pp. 323-330.

Roos, J. E., Chilla, B., Zanetti, M., Schmid, M., Koch, P., Pfirrmann, C. W. and Hodler, J. (2006) 'MRI of meniscal lesions: soft-copy (PACS) and hard-copy evaluation versus reviewer experience', *AJR,* 186(3), pp. 786-790.

Rowland, S., (1896) 'Report on the Application of the New Photography to Medicine and Surgery*, BMJ,* 1(1834), pp. 492.

Royal College of Radiologists (2010), 'Medical image interpretation by radiographers. Guidance for radiologists and healthcare providers'. *Board of the Faculty of Clinical Radiology*, Royal College of Radiologists, London.

The Royal College of Radiologists and the Society and College of Radiographers (2012) 'Team working in clinical imaging'. London: *The Royal College of Radiologists and the Society and College of Radiographers*, London.

Samuel, S., Kundel, H. L., Nodine, C. F. and Toto, L. C. (1995) 'Mechanism of satisfaction of search: eye position recordings in the reading of chest radiographs', *Radiology,* 194(3), pp. 895-902.

Saxton, H., (1992) 'Should radiologists report on every film?', *Clinical Radiology,* 45(1), pp. 1-3.

Scally, A. and Brealey, S. (2003) 'Confidence intervals and sample size calculations for studies of film-reading performance', *Clinical Radiology,* 58(3), pp. 238-246.

Institute of Medicine (1977) 'C. T. Scanning, A policy statement'*. Washington, DC, National Academy of Sciences,*

Schuster, A., (1896) 'On the New Kind of Radiation', *BMJ,* 1(1829), pp. 172-173.

Sim, J. and Wright, C. C. (2005) 'The kappa statistic in reliability studies: use, interpretation, and sample size requirements', *Physical Therapy,* 85(3), pp. 257-268.

Smith, S. and Reeves, P. (2009) 'The extension of the role of the diagnostic radiographer in the UK National Health Service over the period 1995–2009', *European Journal of Radiography,* 1(4), pp. 108-114.

Smith, J. F., (1904) 'XII. The Rontgen-Ray Diagnosis of Renal Calculus', *Annals of Surgery,* 39(5), pp. 748-754.

The College of Radiographers (1997) 'Reporting by Radiographers: a Vision Paper'. *The College of Radiographers:* London.

Society and College of Radiographers (2006) 'Medical Image Interpretation and Clinical Reporting by Non-Radiologists: The Role of the Radiographer. *Society and College of Radiographers:* London

Society and College of Radiographers (2010) 'Medical image interpretation by radiographers. Definitive Guidance'. *Society and College of Radiographers:* London

Society and College of Radiographers (2013) 'Preliminary Clinical Evaluation and Clinical Reporting by Radiographers: Policy and Practice Guidance'. Available at: https://www.sor.org/learning/document-library/preliminary-clinical-evaluation-and-clinical-reporting-radiographers-policy-and-practice-guidance (Accessed 8/7/2013)

Sonin, A. H., Pensy, R. A., Mulligan, M. E. and Hatem, S. (2002) 'Grading articular cartilage of the knee using fast spin-echo proton density-weighted MR imaging without fat suppression', *AJR,* 179(5), pp. 1159-1166.

Starr, S. J., Metz, C. E., Lusted, L. B. and Goodenough, D. J. (1975) 'Visual Detection and Localization of Radiographic Images 1', *Radiology,* 116(3), pp. 533-538.

Swensson, R. G., (1996) 'Unified measurement of observer performance in detecting and localizing target objects on images', *Medical Physics,* 23(10), pp. 1709-1725.

Swinburne, K., (1971) 'Pattern recognition for radiographers', *The Lancet,* 297(7699), pp. 589-590.

Umans, H., Wimpfheimer, O., Haramati, N., Applbaum, Y. H., Adler, M. and Bosco, J. (1995) 'Diagnosis of partial tears of the anterior cruciate ligament of the knee: value of MR imaging', *AJR,* 165(4), pp. 893-897.

University of Hertfordshire and The Institute for Employment Studies (2008). ‘Scope of Radiographic Practice 2008: A report compiled by the University of Hertfordshire in collaboration with the Institute for Employment Studies for the Society and College of Radiographers’. *The Society and College of Radiographers*: London.

van Rijn, J., Klemetso, N., Reitsma, J., Bossuyt, P., Hulsmans, F., Peul, W., den Heeten, G., Stam, J. and Majoie, C. (2006) 'Observer variation in the evaluation of lumbar herniated discs and root compression: spiral CT compared with MRI', *BJR,* 79(941), pp. 372-377.

van Rijn, J. C., Klemetsö, N., Reitsma, J. B., Majoie, C. B., Hulsmans, F. J., Peul, W. C., Stam, J., Bossuyt, P. M. and den Heeten, G. J. (2005) 'Observer variation in MRI evaluation of patients suspected of lumbar disk herniation', *AJR,* 184(1), pp. 299-303.

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C. and Vandenbroucke, J. P. (2008) 'The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies', *Journal of Clinical Epidemiology,* 61pp. 344e349.

Ween, B., Kristoffersen, D. T., Hamilton, G. A. and Olsen, D. R. (2005) 'Image quality preferences among radiographers and radiologists. A conjoint analysis', *Radiography,* 11(3), pp. 191-197.

Weinstein, S., Obuchowski, N. A., Lieber, M. L., Weinstein, S., Obuchowski, N. A. and Lieber, M. L. (2005) 'Clinical evaluation of diagnostic tests', *AJR,* 184(1), pp. 14.

White, L. M., Schweitzer, M. E., Deely, D. M. and Morrison, W. B. (1997) 'The effect of training and experience on the magnetic resonance imaging interpretation of meniscal tears', *Arthroscopy: The Journal of Arthroscopic & Related Surgery,* 13(2), pp. 224-228.

Whiting Penny, Rutjes Anne, Reitsma Johannes, Bossuyt Patrick, , Kleijnen Jos, Penny, W., Anne, R., Johannes, R., Patrick, B. and Jos, K. (2003) 'The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews', *BMC Medical Research Methodology,* 3(1), pp. 25.

Williams, I., (2006) 'Professional role extension for radiographers: opinion article', *South African Radiographer,* 44(2), pp. p. 14-17.

Williams, I. J., (2013) 'Appendicular skeleton: ABCs image interpretation search strategy: peer reviewed article of interest', *South African Radiographer,* 51(2), pp. 9-14.

Wilson, J. (1995). ‘Radiographers reporting plain films—a project report’. *Röntgen Centenary Congress Programme and Abstracts,* Birmingham: p. 887-888.

Wivell, G., Denton, E., Eve, C., Inglis, J. and Harvey, I. (2003) 'Can radiographers read screening mammograms?', *Clinical Radiology,* 58(1), pp. 63-67.

Woznitza, N. and Piper, K. (2011) ‘Can reporting radiographers provide clinically relevant reports?’ *Conference Presentation, Annual Scientific Meeting of Medical Imaging and Radiation Therapy,* Adelaide: Conference Abstract, p.170

Woznitza, N. and Piper, K. (2012a) 'Can reporting radiographers provide clinically relevant reports?' *Conference Presentation, UK Radiology Congress, Manchester: Congress Guide, p.103*

Woznitza, N. (2012b) UK Clinical Research Network Study Portfolio *'Establishing the diagnostic accuracy of radiographer chest x-ray reports and their influence on clinicians' clinical reasoning: A comparison with consultant radiologists'* http://public.ukcrn.org.uk/Search/StudyDetail.aspx?StudyID=14480 (Accessed: 15/04/2014)

Woznitza, N., Burke, S., Amin, S., Patel, K., Grayson, K. and Piper, K. (2013) 'Disagreement in chest x-ray interpretation: comparative analysis between consultant radiologists and a reporting radiographer', *Conference Presentation, UK Radiology Congress,* Liverpool: Conference Abstracts, p.65

Woznitza, N., Piper, K., Burke, S., Patel, K., Amin, S., Grayson, K. and Bothamley, G. (2014a) 'Adult chest radiograph reporting by radiographers: Preliminary data from an in-house audit programme', *Radiography,* 20(3), pp. 223-229.

Woznitza, N., Piper, K., Rowe, S. and West, C. (2014b) 'Optimizing patient care in radiology through team-working: A case study from the United Kingdom', *Radiography,* 20(3), pp. 258-263.

Woznitza, N., (2014c) 'Radiographer reporting', *Journal of Medical Radiation Sciences,* 61(2), pp. 66-68.

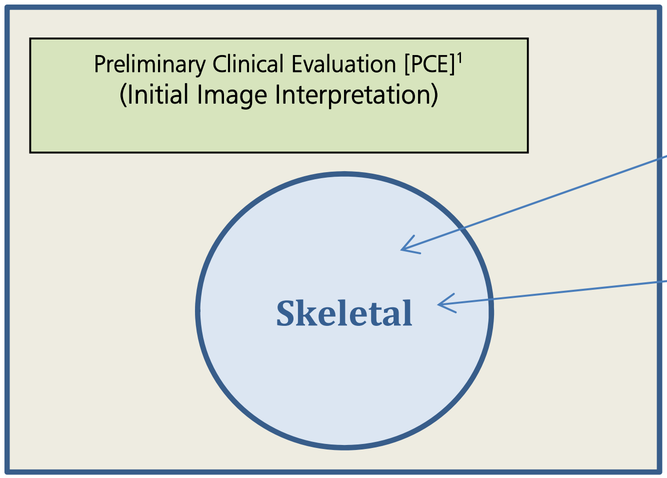
Yatake, H., Takeda, Y., Katsuda, T., Gotanda, R., Yamazaki, H. and Kuroda, C. (2009) 'Film-reading ability of radiographers in detecting gastric cancer during screening using X-ray examination', *Japanese Journal of Radiology,* 27(8), pp. 291-296.

Yerushalmy, J., Harkness, J., Cope, J. and Kennedy, B. (1950) 'The Role of Dual Reading in Mass Radiography.', *American Review of Tuberculosis and Pulmonary Diseases,* 61(4), pp. 443-464.

Yerushalmy, J., (1955) 'Reliability of chest radiography in the diagnosis of pulmonary lesions', *The American Journal of Surgery,* 89(1), pp. 231-240.

Yielder, J., (2014) 'Creating our future: conformity or change?', *Journal of Medical Radiation Sciences,* 61(2), pp. 63-65.

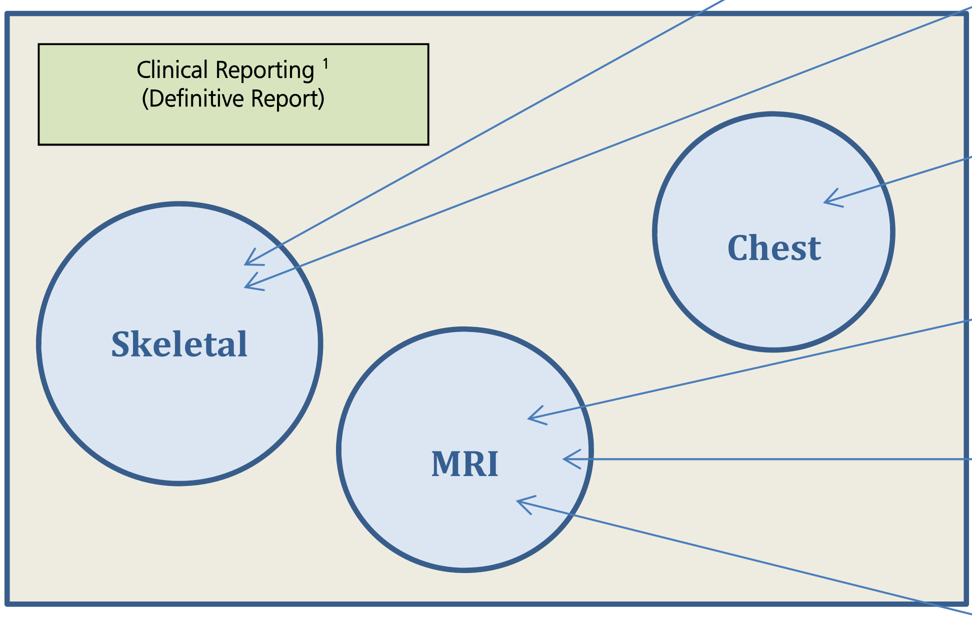
Appendix 1 Research concept map



Study 1 (2009): Improvements in PCE performance were demonstrated after training, in two groups (radiographers and nurses). Differences in performance between the two groups remained, with the radiographer group demonstrating higher accuracy scores.

Study 2 (2009): The accuracy scores and AUC values achieved by the radiographers were statistically higher than those demonstrated by the nurse practitioners and/or casualty officers. The results suggested that radiographers have the ability to formally utilise their knowledge in image interpretation by providing the A/E department with written initial interpretations to assist in the radiographic diagnosis and replace the ambiguous 'red dot' system used to highlight abnormal radiographs.

Study 3 (1999): This clinical study, in an NHS funded multi-centre implementation project, provided evidence that radiographers were able to provide definitive reports on A/E plain film examinations of the musculo-skeletal system to a very high standard.



Study 4 (2005): This study also provided evidence, in an academic setting, that radiographers were able to report on A/E plain film examinations of the musculo-skeletal system to a very high standard. Additionally it demonstrated that, in terms of overall accuracy between reports on A/E and non-A/E referrals, any differences were negligible.

Study 5 (2014): In an academic setting the OSE results for six cohorts of radiographers' chest reports were 95.4%, 95.9% and 89% in terms of sensitivity, specificity and agreement, respectively. The most common errors related to rib appearances or heart size.

Study 6 (2008): In this study a sample of lumbar spine, knee and IAMs MRI examinations were reported by groups of radiologists as part of the construction of an OSE to be used to assess radiographers' competence. Kappa agreement ranged from 0.3 to 0.79, for the lumbar spine and knee examinations and was 1.0 for the IAM cases.

Study 7 (2010): In the OSE the sensitivity, specificity and agreement rates for three cohorts (combined) of radiographers were 99.0%, 99.0% and 89.2%, respectively. These results suggested, that in an academic setting, these groups of radiographers had the ability to correctly identify normal investigations and were able to provide reports on the abnormal appearances to a high standard.

Study 8 (2013): In a pre-implementation study MRI radiographers with postgraduate education and training reported in clinical practice conditions on specific MRI examinations of the knee and lumbar spine to a level of agreement comparable with non-musculoskeletal consultant radiologists.

[1] The Royal College of Radiologists and the Society and College of Radiographers. *Team working in clinical imaging*. London: RCR and SCoR (2012)

Appendix 2 Published papers/report and relative contributions

The following are included in Annex 1

Study 1
Piper, K. J. and Paterson, A. (2009)
'Initial image interpretation of appendicular skeletal radiographs: a comparison between nurses and radiographers', *Radiography,* 15(1), pp. 40-48.

Study 2
Coleman, L. and Piper, K. (2009)
'Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers', *Radiography,* 15(3), pp. 196-202.

Study 3
Piper, K. Paterson, A., and Ryan, C. (1999)
'The implementation of a radiographic reporting service for trauma examinations of the skeletal system in 4 NHS trusts' NHS Executive South Thames funded research project. Canterbury Christ Church University (then College).

Study 4
Piper, K. J., Paterson, A. M. and Godfrey, R. C. (2005)
'Accuracy of radiographers' reports in the interpretation of radiographic examinations of the skeletal system: a review of 6796 cases', *Radiography,* 11(1), pp. 27-34.

Study 5
Piper, K., Cox, S., Paterson, A., Thomas, A., Thomas, N., Jeyagopal, N. and Woznitza, N. (2014)
'Chest reporting by radiographers: Findings of an accredited postgraduate programme', *Radiography,* 20(2), pp. 94-99.

Study 6
Piper, K. and Buscall, K. (2008)
'MRI reporting by radiographers: The construction of an objective structured examination', *Radiography,* 14(2), pp. 78-89.

Study 7
Piper, K., Buscall, K. and Thomas, N. (2010)
'MRI reporting by radiographers: Findings of an accredited postgraduate programme', *Radiography,* 16(2), pp. 136-142.

Study 8

Brealey, S., Piper, K., King, D., Bland, M., Caddick, J., Campbell, P., Gibbon, A., Highland, A., Jenkins, N., Petty, D. and Warren, D. (2013)
'Observer agreement in the reporting of knee and lumbar spine magnetic resonance (MR) imaging examinations: Selectively trained MR radiographers and consultant radiologists compared with an index radiologist',
European Journal of Radiology, 82(10), pp. e597-e605.

Relative contributions

Study 1was a quasi-experimental study which utilised a comparison group pre-post intervention design and was conducted in the academic setting. Images used in a previous study, which compared nurses and junior doctor in A/E radiograph interpretation (Meek et al, 1998) were loaned to me for this study which investigated the effect of training (a short course) and compared a group of radiographers and a group of nurses. The results were analysed using Alternate Free Response Receiver Operating Characteristics (AFROC) methodology, which at the time had not been used for an initial interpretation study of this nature. I led on study design and concepts, was the guarantor of integrity of the study, conducted the literature review, obtained research governance approval, was responsible for data acquisition, analysed and interpreted the data, led on manuscript preparation and editing, gave final approval of version to be submitted for publication and was the corresponding author. The main findings were that improvements were demonstrated after training in both groups (radiographers and nurses), although differences in performance between the two groups remained, with the radiographer group achieving a better overall performance than the nurse group. As patients in MIUs and A/E departments receive treatment based on the initial interpretation of their imaging investigations, by either nurses or radiographers, the improvement after training was encouraging.

Study 2 developed this concept, was also a quasi-experimental study which used a post-test comparison of three groups: junior doctors, nurses and radiographers and was conducted in a clinical environment. I was joint lead on study design and concepts, supervised literature review, and contributed to analysis and interpretation of the data; manuscript preparation and editing. The scores and values achieved by the radiographers were statistically higher than those demonstrated by the participating nurse practitioners and/or casualty officers. The results of the study suggested that radiographers have the ability to formally utilise their knowledge in image interpretation by providing the A/E Dept. with a written initial interpretation to assist in the radiographic diagnosis and therefore replace the ambiguous ‘red dot’ system used to highlight abnormal radiographs. Studies 1 and 2 contributed significantly to the evidence base which resulted in the College of Radiographers advocating the replacement of the ‘red dot’ system with a written initial image interpretation (Society and College of Radiographers, 2010).

Study 3 was a multi-centre observational study which investigated the implementation of a radiographic reporting service in five clinical centres in England. Little research was available at that time, or to date, which has examined this area in this level of detail or with a comparable sample size. The study, which was funded by South Thames NHSE, provided evidence that radiographers were able to report on A/E plain film examinations of the musculo-skeletal system to a very high standard. In general, the speed of availability of reports and the volume reported improved; and the users of the service (A/E Consultants) were extremely or very satisfied with the quality of reports produced. I was overall Project Leader, and specifically was joint lead on study design and concepts, was guarantor of the integrity of the study, jointly conducted the literature review, obtained research ethics approval, jointly collected and analysed all quantitative and qualitative data, was joint lead for project report and editing, and was the corresponding author.

Study 4 was an observational study which compared the diagnostic performance of radiographers (three cohorts) in the interpretation of appendicular skeleton and axial skeleton radiographs (n=6796), in an OSE at the end of an accredited postgraduate programme. Little research was available at this time, or to date, which has examined this area in detail or with a comparable sample size. The study provided evidence that radiographers were able to report on A/E plain film examinations of the musculo-skeletal system to a very high standard. It also demonstrated that, in terms of overall accuracy between reports on A/E and non-A/E referrals, any differences were negligible. I led on study design and concepts, was guarantor of the integrity of the study, conducted the literature review, obtained research governance approval, collected data, inputted all quantitative data, jointly analysed data, interpreted and analysed the data, was lead for manuscript preparation and editing, approved the final version submitted for publication, and was the corresponding author.

Study 5 investigated the reporting of plain radiographs of the chest which is an area undertaken by radiographers, but to a lesser extent than skeletal reporting. The study analysed the results of six cohorts of radiographers who completed a 100 station OSE at the end of an accredited postgraduate programme, in an academic setting and in a controlled environment. The OSE results for six cohorts of radiographers’ chest reports (n=4000) were 95.4%, 95.9% and 89% in terms of sensitivity, specificity and agreement, respectively. The most common errors were related to rib appearances or heart size; and based on a detailed literature review included in the article, the types of errors made by the radiographers were likely to be similar to those made in the clinical setting by consultant radiologists of varying experience. I led on study design and concepts, was guarantor of the integrity of the study, conducted the literature review, obtained research governance approval, was responsible for data acquisition, analysed and interpreted the data, led on manuscript preparation and editing, gave final approval for the version to be submitted for publication and was the corresponding author.

Study 6 examined the process of constructing an objective structured examination (OSE) and as part of the process, in an observer agreement study, compared MRI reports produced by consultant radiologists for a number of different anatomical structures of the lumbar/thoracic spine, IAM and the knee. I led on study design and concepts, was guarantor of the integrity of the study, conducted the literature review, obtained research governance approval, was responsible for data acquisition, analysed and interpreted the data, led on manuscript preparation and editing, gave final approval for version to be submitted for publication and was the corresponding author. When analysed using Kappa, agreement ranged from 0.3 to 0.79, for the lumbar spine and knee examinations. With the exception of one knee study (Bryan et al, 2001) this was the first study that reported on observer agreement between radiologists, in the UK, in the interpretation of MRI investigations.

Study 7 was an observational study which analysed the OSE results for the first three cohorts of MRI radiographers who completed the OSE. Agreement between a representative sample of the radiographers and the radiologists' reports was also investigated. I led on study design and concepts, was guarantor of the integrity of the study, conducted the literature review, obtained research governance approval, was responsible for data acquisition, analysed and interpreted the data, led on manuscript preparation and editing, gave final approval for version to be submitted for publication and was the corresponding author. The sensitivity, specificity and agreement rates for three cohorts (combined) of radiographers were 99.0%, 99.0% and 89.2%, respectively. The levels of agreement were similar, when the Kappa values for the groups of radiographers and radiologists were compared. These results suggested therefore that in an academic setting, these groups of radiographers had the ability to correctly identify normal investigations and were able to provide a report on the abnormal appearances to a high standard.

Study 8 was an observer agreement study which also examined the potential implementation into practice and the impact on patient management. I contributed to the study concepts and design; and manuscript editing and review. In a pre-implementation study, selected MRI radiographers with postgraduate education and training, reported in clinical practice conditions, on specific MRI examinations of the knee and lumbar spine, to a level of agreement comparable with non-musculoskeletal consultant radiologists.

Appendix 3 Quality criteria developed by Brealey (2004)

PART 1: Study eligibility and design

A Study eligibility:

A1 *Inclusion criteria*

For a study to be eligible for inclusion it must satisfy the criteria below:

- Radiographer(s) were compared with a reference standard to assess their plain radiograph reading performance
- Must include or have the potential to calculate an appropriate statistic that reflects accuracy (e.g. sensitivity, specificity).

A2 *Exclusion criteria*

A study will be excluded if:

- Images from other modalities (e.g. mammograms, ultrasound scans)
- Not performed during 1971-2002/10
- Case reports
- Visual search strategy study
- Duplication of data

A3 Is the study eligible (please explain why below)?

- Yes
- No

B Study design:

B1 In what setting was the study conducted?

- outside of routine clinical practice e.g. postgraduate course (which will be a study of the efficacy of the film reading performance of radiographers).
- during routine clinical practice (which will be a study of the effectiveness of the film reading performance of radiographers).
-

B2 What was the design of the study as an assessment of the film reading performance of a cohort(s) of observers?

- Cohort A *versus* reference standard: How accurate is cohort A when interpreting plain films?
- Cohort A *versus* Cohort B *versus* reference standard: How accurate is cohort A when interpreting plain films? How accurate is cohort B when interpreting plain films? How does cohort A compare to cohort B when interpreting plain films?
- Cohort A *versus* Cohort B *versus* Cohort C *versus* reference standard: How accurate is cohort A when interpreting plain films? How accurate is cohort B when interpreting plain films? How accurate is cohort C when interpreting plain films? Is there any difference in performance between the cohorts studied?

B3 What was the design of the study as described below?

- *diagnostic accuracy*: to assess the film reading performance of one (or more) group of observers in controlled (ideal) conditions.
- *diagnostic performance*: to assess the film reading performance of one group of observers during clinical practice.
- *diagnostic outcome*: to assess the film reading performance of two (or more) group of observers during clinical practice

B4 What was the focus of the study with regards to the role of the observers being evaluated?

- *Pattern recognition study:* recognition of the presence of an abnormality (e.g. red dot system); or
- *Reporting study*: ability to produce a precise diagnosis (e.g. correct abnormality and location) using a combination of codes and free text.
- other (specify below).

PART 2: Quality criteria checklist

The quality criteria checklist has been subdivided into two sections: identification of biases and general methodological factors.

Section 1: Identification of biases

Each criterion is scored as:

- DONE (A) - there is evidence from an (un)published report or via personal communication that the criterion was achieved.
- NOT CLEAR (B) - if there is insufficient information from an (un)published report or via personal communication that the criterion was achieved. Missing information will be sought by the main reviewer.
- NOT DONE (C) - there is evidence from an (un)published report that the criterion was not achieved; or there is evidence from personal communication that the criterion was not achieved.
- Not applicable (N/A) - the criterion that the question is addressing is clearly not relevant to the particular study.

Can you please record the score you chose for each criterion by ticking the relevant box. Please record why you chose that score for each criterion under *Comment*:

Subjects   [external validity]

If the study was conducted outside of routine clinical practice then answer section A. If the study was conducted during routine clinical practice then answer section B. Answer Section C for all studies to judge whether observers were appropriately selected.

Film selection

A Studies conducted outside of clinical practice (film cohort bias: spectrum; film filtering: eligibility criteria)

A1 Is spectrum bias present?

- ❑ Score DONE (A) if an attempt was made to include a non-random case mix based on at least three of the following factors: prevalence of disease, severity of disease, range of disease type, pertinent areas of the body; or
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if:
- there was no record of the case mix of the films
- two or less factors were taken into consideration when generating the case mix.
- ❑ N/A.

Comment:

A2 Are specific eligibility criteria stated for those included / excluded (*film filtering bias*)?

- ❑ Score DONE (A) if criteria are reported for all those films that were eligible for inclusion or exclusion from the study and the total number of films included is given as well as the number included/ excluded.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if criteria or numbers are not reported.
- ❑ N/A.

Comment:

B Studies conducted during clinical practice (referral biases: centripetal, popularity; film cohort: population; film filtering: eligibility criteria, film selection)

Questions B1-2 provide only information. A judgement from this information is required to assess the presence or absence of these referral biases.

B1 Is the establishment(s) where the study was undertaken stated (*centripetal bias*)?

- ❑ Score DONE (A) if the establishment is the place of origin of the study.
- ❑ Score NOT DONE (C) if not reported.

Comment:

B2 Is the establishment from where the patients were referred stated (*popularity bias*)?

- ❑ Score DONE (A) if the establishment is clearly stated e.g. A&E department.
- ❑ Score NOT DONE (C) if not reported.

Comment:

B3 Is population bias present?

- ❑ Score DONE (A) if:
- a series of films over a suitable time period was included; <u>or</u>
- a valid random sample of films were selected in a way so that the professionals

responsible for interpreting the films had no choice as to what films they interpreted and the random process is described explicitly, e.g. the use of random number tables.

- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if:
- there is no statement as to the length of the time period during which the consecutive series of films were interpreted; or the series of films that were included was not during a long enough time period.
- the allocation procedure for randomisation is not described; or alternation such as reference to case record numbers, dates of birth, day of the week or any other such approach was used in the selection of films.
- ❑ N/A.

Comment :

B4 Are specific eligibility criteria stated for those included / excluded?

- ❑ Score DONE (A) if criteria are reported for all those films that were eligible for inclusion or exclusion from the study and the total number of films included is given as well as the number included/ excluded; or is it clear that all films were included.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if criteria or numbers are not reported.
- ❑ N/A.

Comment:

B5 Is film selection bias present?

- ❑ Score DONE (A) if:
- all films eligible to be included in the study were interpreted by the observers under evaluation; and
- observers could not choose which eligible films to interpret.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if:
- not all the eligible films were interpreted by the observers; or
- the observers could choose which eligible films to interpret (i.e. systematic exclusions).
- ❑ N/A.

Comment:

Observer selection

C Relevant to all studies

C1 Is observer cohort bias present?

- ❑ Score DONE (A) if an appropriate group of observers were selected.
- ❑ Score NOT CLEAR (B) if there is insufficient information.

- ❑ Score NOT DONE (C) if an inappropriate group of observers were selected.
- ❑ N/A.

Comment:

C2 Is observer cohort comparator bias present?

- ❑ Score DONE (A) if the study group (received training) and control group (no training) were matched according to the following characteristics: professional group; number of years experience in the profession; number of years experience in a relevant speciality (e.g. A&E); number of years experience interpreting images (e.g. ultrasound).
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the study group and control group were not matched according to the above characteristics.
- ❑ N/A.

Comment:

Study [internal validity]

All studies should be assessed in relation to the following criteria:

D Application of the reference standard

D1 Is verification bias present?

- ❑ Score DONE (A) if all the films interpreted by the observers under evaluation were also interpreted by the reference standard or a correction is performed by the authors.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if not all films interpreted by the observers under evaluation were also interpreted by the reference standard.
- ❑ N/A.

Comment:

D2 Is work-up bias present?

- ❑ Score DONE (A) if the interpretation made by the observers under evaluation is not used to decide whether the reference standard is applied or a correction is performed by the authors.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the interpretation made by the observers under evaluation is used to decide whether the reference standard is applied.
- ❑ N/A.

Comment:

D3 Is incorporation bias present?

- ❑ Score DONE (A) if the interpretation of an observer under evaluation is not incorporated into the evidence used to diagnose the disease or is itself not used as the reference standard.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the interpretation of an observer under evaluation is incorporated into the evidence used to diagnose the disease or is itself used as the

reference standard.
- N/A.

Comment:

E **Measurement of results** (disease progression; withdrawal bias: indeterminate observer interpretations, follow-up; observer variability: inter-observer; intra-observer; arbiter variability: inter-arbiter, intra-arbiter).

E1 Is disease progression bias present?

- Score DONE (A) if appropriate radiological and clinical review is used.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if inappropriate clinical and radiological review is used.
- N/A.

Comment:

E2 Are there any indeterminate (i.e. equivocal, non-diagnostic) observer interpretations?

- Score DONE (A) if all films and subsequent interpretations are included irrespective of their indeterminability.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if films are excluded due to indeterminate interpretations.
- N/A.

Comment:

E3 Are there any patients lost to follow-up?

- Score DONE (A) if all films and clinical information is available for verification.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if patients are excluded or films not reported owing to loss.
- N/A.

Comment:

E4 Is any attempt made to assess intra-observer variability?

- Score DONE (A) if for a subsample of the films interpreted data are reported statistically, or illustrated in a ROC curve.
- Score NOT CLEAR (B) if there is insufficient information.
- Score NOT DONE (C) if no data are provided.
- N/A.

Comment:

E5 Is any attempt made to assess inter-observer variability?

- Score DONE (A) if for a subsample of the films interpreted data are reported statistically, or illustrated in a ROC curve.
- Score NOT CLEAR (B) if there is insufficient information.

- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

E6 Is any attempt made to assess intra-arbiter variability?

- ❑ Score DONE (A) if for a subsample of the interpretations compared data are reported statistically, or illustrated in a ROC curve.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

E7 Is any attempt made to measure inter-arbiter variability?

- ❑ Score DONE (A) if for a subsample of the interpretations compared data are reported statistically, or illustrated in a ROC curve.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if no data are provided.
- ❑ N/A.

Comment:

F Independence of interpretations

F1 Is observer review bias present?

- ❑ Score DONE (A) if the observers being evaluated were blinded or unaware of the interpretation made by the reference standard when interpreting the films.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the observers being evaluated were aware of the interpretation made by the reference standard when interpreting the films.
- ❑ N/A.

Comment:

F2 Is reference standard review bias present?

- ❑ Score DONE (A) if the reference standard was blinded or unaware of the interpretation made by the observers under evaluation.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the reference standard was aware of the interpretation made by the observers under evaluation.
- ❑ N/A.

Comment:

F3 Is observer bias present?

- ❑ Score DONE (A) if all observers always interpreted the films independent of each other.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if observers did not always interpret the films independent of each other.
- ❑ N/A.

Comment:

F4 Is observer comparator bias present?

- ❑ Score DONE (A) if all observers interpreted the same or a similar set of films independent of each other.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if observers did not always interpret the same or a similar set of films independent of each other.
- ❑ N/A.

Comment:

F5 Is co-image bias present?

- ❑ Score DONE (A) if all observers only had access to the films that they were being asked to interpret and not images from other modalities in relation to the same examination.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if observers had access to images from other modalities in relation to the films that they were being asked to interpret.
- ❑ N/A.

Comment:

F6 Is arbiter review bias present?

- ❑ Score DONE (A) if the arbiter was not one of the observers under evaluation or the reference standard.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the arbiter was one of the observers under evaluation and/ or the reference standard.
- ❑ N/A.

Comment:

F7 Is arbiter bias present?

- ❑ Score DONE (A) if the arbiter was blind or unaware as to whether the report was made by an observer under evaluation or the reference standard.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the arbiter was aware of who was responsible for either of the reports.
- ❑ N/A.

Comment:

F8 Is film access bias present?

- ❑ Score DONE (A) if the arbiter judged whether interpretations agreed or not without access to the films.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the arbiter made use of the films during the process of judging whether interpretations agreed or not.
- ❑ N/A.

Comment:

ADDITIONAL VALIDITY CRITERIA FOR STUDIES COMPARING TWO (OR MORE) COHORTS

F9 Is cohort comparator bias present?

- ❑ Score DONE (A) if the cohorts of observers interpreted the same films independent of each other.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the cohorts of observers did not always interpret the films independent of each other; <u>or</u> did not report on the same films.
- ❑ N/A.

Comment:

F10 Is co-image comparator bias present?

- ❑ Score DONE (A) if both cohort of observers had similar access to the relevant plain films and did not have access to images from other modalities.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if both cohort of observers did not have similar access to the relevant plain films and one cohort of observers had access to images from other modalities.
- ❑ N/A.

Comment:

F11 Is arbiter comparator bias present?

- ❑ Score DONE (A) if the arbiter was blind or unaware as to who was responsible for the interpretations when judging whether they agreed or not.
- ❑ Score NOT CLEAR (B) if there is insufficient information.
- ❑ Score NOT DONE (C) if the arbiter was aware of who was responsible for the interpretations when judging whether they agreed or not.
- ❑ N/A.

Comment:

Section 2: General methodological standards

Each criterion is scored as:

- ❑ DONE (A) - there is evidence from an (un)published report or via personal communication that the criterion was achieved.
- ❑ NOT DONE (C) - there is no evidence from an (un)published report that the criterion was achieved; or there is evidence from personal communication that the criterion was not achieved; or there is no evidence from an (un)published report or via personal communication that the criterion was achieved.
- ❑ Not applicable (N/A) - the criterion that the question is addressing is clearly not relevant to the particular study.

Can you please record the score you chose for each criterion by ticking the relevant box(es). Please record why you chose that score for each criterion under *Comment*:

G Subjects (films)

G1 Was an appropriate sample size considered?

- ❑ Score DONE (A) if the study:
- measured the performance of a single cohort of observers and the sample size was calculated according to how precise an estimate of the sensitivity and specificity was required.
- reports an attempt to calculate the sample size required to detect clinically important effects as statistically significant between two (or more) cohorts of observers, and if possible, record the power under *Comment*.
- ❑ Score NOT DONE (C) if:
- no reference is made to the sample size required.
- no power calculation is stated, or the study did not attempt to calculate the sample size required.
- ❑ N/A.

Comment:

H Study

H1 Was a normal/abnormal report adequately defined?

- ❑ Score DONE (A) if an explicit attempt was made to adequately define a normal/abnormal report.
- ❑ Score NOT DONE (C) if a normal/abnormal report was not adequately defined.
- ❑ N/A.

Comment:

H2 Was the performance of the observers placed in the context of the diagnostic sequence (i.e. referral filters e.g. red dot system, casualty officers [cold], hot)?

- ❑ Score DONE (A) if the study made an explicit attempt to report the process through which the films had passed before they were interpreted by the observers under evaluation.
- ❑ Score NOT DONE (C) if the study did not report the context in which the films were interpreted.
- ❑ N/A.

Comment:

H3 If the combined performance of two (or more) different groups of observers is assessed was the contribution of the individual groups to the overall validity of the combination of groups determined?

- ❑ Score DONE (A) if every single group within a combination of groups was evaluated.
- ❑ Score NOT DONE (C) if not every single group of a combination of groups were evaluated.
- ❑ N/A.

Comment:

H4 Was an appropriate (valid) reference (“gold” or "criterion") standard used?

Score DONE (A) if the study reported a suitable reference standard:

- A1: a double/triple blind consultant radiological report.
- A2: a single consultant radiological report that was validated in an acceptable way e.g. via clinical follow-up.
- A3: a single consultant radiological report that was not validated.
- Score NOT DONE (C) if an inappropriate reference standard is reported e.g. a combination of radiologists at different grades, the observers under evaluation were also used as the reference standard or included in the process of generating the reference standard; not reported in the paper.
- N/A.

Comment:

H5 Was an appropriate (valid) arbiter used?

Score DONE (A) if the study used a suitable arbiter:

- A1: external: panel
- A2: external: consultant radiologist
- A3: internal: panel
- A4: internal: consultant radiologist
- A5: radiographer(s) trained to report and if unsure an independent consultant radiologist.
- A6: untrained radiographer(s) and if unsure an independent consultant radiologist.
- Score NOT DONE (C) if: the study reported an inappropriate arbiter e.g. independent untrained radiographer(s) with no referral to radiologist, a person under evaluation is responsible for comparing the reports; not reported in the paper.
- N/A.
- *Comment:*

H6 Was a control used in the study (*appropriate choice of control activity*)?

- Score DONE (A) if an appropriate control was used within the context of the particular study.
- Score NOT DONE (C) if an inappropriate control was used; or a control was appropriate but not used.
- N/A.

Comment:

I Interpretation

I1 Were films appropriately analysed for pertinent subgroups?

- Score DONE (A) if an attempt was made to analyse the observers performance for pertinent medical subgroups, e.g. areas of the body.
- Score NOT DONE (C) if there was no attempt to analyse pertinent medical subgroups.
- N/A.

Comment:

I2 Was the data presented in enough detail to allow for the calculation of appropriate indices of performance (e.g. sensitivity and specificity) and confidence intervals?

- ❑ Score DONE (A) if the data was presented in enough detail to calculate the above.
- ❑ Score NOT DONE (C) if the data was NOT presented in enough detail to calculate the above.
- ❑ N/A.

Comment:

I3 Are indeterminate observer interpretations appropriately presented?

- ❑ Score DONE (A) if a study reported:
- all of the appropriate positive, negative and indeterminate interpretations; and
- whether indeterminate interpretations had been included or excluded when indices of performance were calculated.
- ❑ Score NOT DONE (C) if the study did not:
- attempt to categorise reports as positive, negative, and indeterminate.
- state whether indeterminate results had been included or excluded when indices of performance were calculated.
- ❑ N/A.

Comment:

Part 2: Quality criteria checklist

Section 1: Identification of biases in the overall design of the study

Subjects [internal and external validity]

Film selection

A Studies conducted outside of clinical practice

A1 *Spectrum bias* - this is present when not all of the following factors are considered when selecting the sample of films: prevalence of disease, severity of disease, disease type, and areas of the body.

A2 *Film filtering bias* - this is present if there is no record of the criteria used to determine which films were eligible for inclusion or exclusion. This bias is also present if the total number of films is not given and the number included/ excluded.

B Studies conducted during clinical practice

B1 *Centripetal bias* - this is present if there is no record of the establishment where the study was undertaken.

B2 *Popularity bias* - this is present if the establishment from where patients were referred is not clearly stated.

B3 *Population bias* - this is present if a series of films included in a sample was not over a suitable time period or was not a valid random sample. The decision as to whether the observers interpreted a series of films over a long enough time period is a subjective one.

B4 *Film filtering bias* - see A2.

B5 *Film selection bias* - this occurs if the observers under evaluation do not interpret all the films that are eligible to included in the study and/ or have the opportunity to choose which eligible films they want to interpret.

Observer selection

C Relevant to all studies

C1 *Observer cohort bias* - this occurs if an inappropriate selection of observers are included in a study with regards to the research question that is being addressed.

C2 *Observer cohort comparator bias* - this occurs if two (or more) groups of observers are compared without the appropriate use of matching. For studies that assess the effectiveness of a training programme and are comparing a study group (receive training) with a control group (no training) the two groups should be matched for the characteristics listed to ensure comparability.

Study [internal validity]

D Application of the reference standard

D1 *Verification bias* - this occurs when not all of the films interpreted by the observers under evaluation are interpreted by the same reference standard for any reason e.g. economic limitations, decisions based on clinical signs and symptoms.

D2 *Work-up bias* - this occurs when not all the films receive definitive confirmation with the reference standard due to the interpretation of the observers under evaluation. Using this definition, if work-up bias is present then verification bias is also present but not *vice versa*.

D3 *Incorporation bias* - this occurs if the report of an observer under evaluation is incorporated into the evidence used to diagnose the disease. This also occurs if the report of the observer under evaluation is used as the reference standard e.g. the report of an observer under evaluation within a cohort, such as a radiologist, is used as the reference standard. Incorporation bias is not present if the study is designed to follow the progression of a disease, and a definitive endpoint reference standard is used for diagnosis.

E Measurement of results

E1 *Disease progression bias* - this occurs if there is a long time period between the initial report and subsequent clinical follow up. If the reference standard only involves reporting films then this bias is not applicable. However, if the reference standard includes clinical follow-up, it is important that there is appropriate radiological review. This is to ensure that the initial film, for example, was incorrectly interpreted by an observer because of a missed overt fracture rather than the film being correctly reported but an occult fracture resulted in the patient re-attending.

E2 *Indeterminate interpretation bias* - this is present if not all indeterminate interpretations are included when measuring observers performance. If films are excluded for this reason prior to the application of the reference standard this will introduce work-up bias.

E3 *Loss to follow-up bias* - this occurs if information is systematically lost so that the reference standard can not be applied.

E4 *Intra-observer variability bias* - this occurs if the observers under evaluation did not re-interpret a subsample of the films to measure their consistency in the interpretation of films.

E5 *Inter-observer variability bias* - this occurs if the observers within a cohort did not report on the same subsample of films. If only one observer this is not applicable.

E6 *Intra-arbiter variability bias* - this occurs if the same arbiter did not re-apply the criteria used to judge whether there is concordance between interpretations on a subsample of cases.

E7 *Inter-arbiter variability bias* - this occurs if two independent arbiters did not compare a subsample of the observer interpretations with the reference standard to assess whether the criteria was applied consistently by different people

F Independence of interpretations

F1 *Observer review bias* - this occurs if the observers being evaluated are aware of the interpretation made by the reference standard when interpreting the films. If the reference standard used is clinical follow-up, so long as it is not a retrospective study the results of the definitive diagnosis must be unknown at the time of the interpretation by the observers under evaluation. Thus, the bias is absent.

F2 *Reference standard review bias* - this occurs if the interpretations of the observers under evaluation are known when the diagnosis is made by the reference standard.

F3 *Observer bias* - this occurs if the individual observers within a cohort do not interpret the films independent of each other.

F4 *Observer comparator bias* - this occurs if an attempt is made to compare the performance of observers within a cohort and not all observers interpreted the same or a similar set of films independent of each other.

F5 *Co-image bias* - this occurs if additional images were available to a cohort of observers other than those they were being assessed to interpret with the exception of previous plain films.

F6 *Arbiter review bias* - this occurs if the arbiter was one of the observers under evaluation or was the reference standard.

F7 *Arbiter bias* - this occurs if the arbiter was aware as to whether the interpretation was made by the observer(s) under evaluation or the reference standard.

F8 *Film access bias* - this occurs if the arbiter had access to films whilst judging whether interpretations agreed or not. Their interpretation can incorrectly influence the decision as to whether the reports agree or not, or as to which report is correct.

ADDITIONAL VALIDITY CRITERIA FOR STUDIES COMPARING TWO (OR MORE) COHORTS

F9 *Cohort comparator bias* - this occurs if the cohorts of observers did not interpret the same films independent of each other. For example, a study may have compared radiographers performance with the reference standard and radiologists performance with the reference standard. Both the cohort of radiographers and radiologists should interpret the films independently. Furthermore, the two cohorts should report on the same or a comparable batch of films.

F10 *Co-image comparator bias* - this occurs if one cohort of observers had access to images from other modalities.

F11 *Arbiter comparator bias* - this occurs if the arbiter was aware as to which of the interpretations was made by the different cohort of observers.

Section 2: General methodological factors

G Subjects (films)

G1 (a) If the study is measuring the performance of a single cohort of observers the sample size should be calculated according to how precise an estimate of the sensitivity and specificity is required. (b) Studies comparing cohorts should make use of a power calculation.

H Study

H1 Whether the definition of normal or abnormal is acceptable is a subjective one. The important issue is whether a definition was available.

H2 It is important that a study describes the diagnostic sequence through which films pass as this will affect the case mix of films that the observers interpret and subsequently the generalisability of the results. This criterion will not be applicable to postgraduate studies.

H3 Some studies may assess the combined performance of two groups of observers such as the interpretation made by a nurse practitioner having seen the interpretation made by a radiographer. This type of study should also assess the performance of the two groups separately to identify the contribution of each group to the combined effort.

H6 The relevant control may vary, if one is necessary, depending on the research question.

Appendix 4 Completed checklist – Study 1

Piper, K. J., & Paterson, A. (2005). Initial image interpretation of appendicular skeletal radiographs: a comparison between nurses and radiographers (2005) *Radiography*, *15*(1), 40-48.

A quasi-experimental study which utilised a comparison group pre-post intervention design and was conducted in the academic setting. The study investigated the effect of training (a short course) and compared a group of radiographers and a group of nurses. The results were analysed using Alternate Free Response Receiver Operating Characteristics (AFROC) methodology

Study assessed using Diagnostic accuracy criteria (Appendix 3)

Criteria	Score	Comments
A1	A	
A2	A	
C1	A	
C2	A	
D1	A	
D2	N/A	
D3	A	
E1	N/A	
E2	A	
E3	N/A	
E4	C	Intra-observer variability not assessed
E5	A	
E6	C	Intra-arbiter variability not assessed
E7	C	Single arbiter
F1	A	
F2	A	
F3	A	
F4	A	
F5	A	
F6	A	
F7	C	Arbiter was aware of whether report was by the reference standard or an observer under evaluation
F8	A	
F9	A	
F10	A	
F11	A	
G1	A	
H1	A	
H2	N/A	
H3	N/A	
H4	A2	
H5	A3	
H6	N/A	
I1	N/A	
I2	A	
I3	A	

Appendix 5 Completed checklist – Study 2

Coleman, L., & Piper, K. (2009). Radiographic interpretation of the appendicular skeleton: A comparison between casualty officers, nurse practitioners and radiographers. *Radiography*, *15*(3), 196-202

A quasi-experimental study which utilised a comparison pre-post intervention design and was conducted in the clinical setting. The study investigated the image interpretation performance and confidence of three groups: radiographers, nurses and casualty officers. The results were analysed using Alternate Free Response Receiver Operating Characteristics (AFROC) methodology.

Study assessed using Diagnostic accuracy criteria (Appendix 3)

Criteria	Score	Comments
A1	A	
A2	A	
C1	A	
C2	A	
D1	A	
D2	N/A	
D3	A	
E1	N/A	
E2	A	
E3	N/A	
E4	C	Intra-observer variability not assessed
E5	A	
E6	B	Intra-arbiter variability included for 10% of cases. Agreement values not included.
E7	C	Single arbiter – see E6
F1	A	
F2	A	
F3	A	
F4	A	
F5	A	
F6	A	
F7	C	Arbiter was aware of whether report was by the reference standard or an observer under evaluation
F8	A	
F9	A	
F10	A	
F11	A	
G1	A	
H1	A	
H2	N/A	
H3	A	
H4	A1	Consensus of 3 experienced plain film reporters
H5	A3	
H6	N/A	
I1	N/A	
I2	A	
I3	A	

Appendix 6 Completed checklist – Study 3 & 4

From Brealey (2004)

Please note amendments/additions by Piper for this commentary are highlighted for clarity

Table A2.3.1 Results from quality criteria checklist for diagnostic accuracy studies

	* Henderson 1999[14]	*Wilson 1999[22]	Piper & Paterson 1997[17]	*Callaway et al 1997[3]	Piper et al 2005	Study 4
A1	A	A	A	A	A	
A2	C	A	A	A	A	
C1	A	A	A	A	A	
C2	N/A	A	A	N/A	N/A	
D1	A	A	A	A	A	
D2	A	A	A	A	N/A	
D3	A	A	A	A	A	
E1	N/A	N/A	N/A	N/A	N/A	
E2	A	A	A	A	A	
E3	N/A	N/A	N/A	N/A	N/A	
E4	C	C	C	C	N/A	
E5	C	C	C	C	A	
E6	C	C	C	C	C	
E7	C	C	A	C	N/A	
F1	A	A	A	A	A	
F2	A	A	A	A	A	
F3	A	A	A	A	A	
F4	A	N/A	N/A	N/A	A	
F5	A	A	A	A	A	
F6	A	A	A	A	A	
F7	B	B	C	C	C	
F8	C	C	C	C	A	
F9	N/A	A	A	A	N/A	
F10	N/A	A	A	A	N/A	
F11	N/A	A	C	C	N/A	
G1	C	C	C	C	C	
H1	A	A	A	A	A	
H2	N/A	N/A	N/A	N/A	N/A	
H3	N/A	N/A	N/A	N/A	N/A	
H4	A3	A1	A1	A3	A1	
H5	A6	A6	A3	A3	A2	
H6	N/A	A	N/A	N/A	N/A	
I1	C	A	A	N/A	A	
I2	C	C	A	N/A	A	
I3	C	N/A	N/A	A	N/A	

*Original references included in Brealey (2004) and not included in this submission

H1	A	A	A	A	A	A
H2	N/A	N/A	N/A	N/A	N/A	N/A
H3	N/A	N/A	N/A	N/A	N/A	N/A
H4	A1	A3	A3	A1	A3	A1
H5	A5	A4	A5	C	C	A6
H6	A	N/A	C	N/A	C	N/A
I1	C	N/A	C	A	C	N/A
I2	A	C	C	C	A	N/A
I3	N/A	N/A	N/A	A	C	N/A

*Original references included in Brealey (2004) and not included in this submission

Table A2.3.2 Results from quality criteria checklist for diagnostic performance studies

	Loughran *et al* 1996	*** Raynor 1999[15]**	***Manning 1999[12]**	***Eyres & Williams 1999[9]**	**Piper *et al* 1999 Study 3**
B1	A	A	A	A	A
B2	A	A	A	A	A
B3	A	A	A	A	A
B4	A	A	A	A	A
B5	A	A	A	A	A
C1	A	A	A	A	A
C2	N/A	N/A	N/A	N/A	N/A
D1	A	A	A	A	A
D2	A	A	A	A	A
D3	A	A	A	C	C
E1	N/A	N/A	N/A	N/A	N/A
E2	N/A	N/A	A	A	C
E3	A	A	A	B	A
E4	C	C	C	C	C
E5	C	N/A	C	C	C
E6	C	C	N/A	C	C
E7	C	C	N/A	C	N/A
F1	A	A	C	A	A
F2	C	C	A	A	C
F3	A	A	B	A	A
F4	N/A	N/A	N/A	N/A	N/A
F5	A	A	A	A	A
F6	C	C	C	C	C
F7	C	C	C	C	C
F8	C	C	A	B	C

G1	C	C	C	C	A
H1	A	A	A	A	A
H2	A	A	A	C	A
H3	N/A	N/A	N/A	N/A	N/A
H4	A3	A3	A3	C	C
H5	C	C	C	C	A3
H6	N/A	N/A	N/A	N/A	N/A
I1	C	C	A	A	A
I2	C	C	A	A	A
I3	N/A	N/A	N/A	N/A	A

*Original references included in Brealey (2004) and not included in this submission

Table A2.4.3 Ranking of studies by mean score

Study	Mean Score	Rank
18 *	10.1	1
17	19.3	2
22	24.6	3
10	25.8	4
3	28.0	5
11	30.6	6
26b	31.7	7
28 **	33.2	8
31	34.4	9
30	34.7	10
12	35.2	11
23	36.8	12
27	37.0	13
26a	37.4	14
13	38.4	15
15	39.8	16
29	40.7	17
5	40.8	18
19	40.8	19
7	41.0	20
4	41.8	21
20	42.2	22
9	42.4	23
14	43.7	24
8	46.7	25
6	46.7	26
32	48.2	27
21	49.7	28
24	53.5	29
16	54.7	30
34	55.2	31
2	55.3	32
1	57.5	33
25	61.8	34
33	80.0	35

* Piper et al (2005) – Study 4 (Annex 1) of this submission

** Piper et al (1999) – Study 3 (Annex 1) of this submission

Table A2.4.4 Study ranking using the mean score: diagnostic accuracy studies

Study	Mean Score	Rank
18 *	10.1	1
17	19.3	2
22	24.6	3
10	25.8	4
3	28.0	5
11	30.6	6
23	36.8	12
27	37.0	13
13	38.4	15
14	43.7	24
2	55.3	32
All studies	**31.8**	

* Piper et al (2005) – Study 4 (Annex 1) of this submission

Table A2.4.5 Study ranking using the mean score: diagnostic performance studies

Study	Mean Score	Rank
28 **	33.2	8
30	34.7	10
12	35.2	11
26a	37.4	14
15	39.8	16
29	40.7	17
19	40.8	19
7	41.0	20
9	42.4	23
8	46.7	25
32	48.2	27
21	49.7	28
34	55.2	31
25	61.8	34
33	80.0	35
All studies	**45.8**	

** Piper et al (1999) – Study 3 (Annex 1) of this submission

Appendix 7 Completed checklist – Study 5

Piper. K, Cox. S, Paterson. A, Thomas. A, Thomas. N, Jeyagopal. N, Woznitza. (2014) Chest Reporting by Radiographers: Findings of an accredited postgraduate programme. *Radiography*, 20(2), 94-99

Study assessed using Diagnostic accuracy criteria (Appendix 3)

Criteria	Score	Comments
A1	A	
A2	B	
C1	A	
C2	A	
D1	A	
D2	N/A	
D3	A	
E1	N/A	
E2	A	
E3	N/A	
E4	C	Intra-observer variability not assessed
E5	A	
E6	B	Intra-arbiter variability not assessed
E7	C	Single arbiter – see E6
F1	A	
F2	A	
F3	A	
F4	A	
F5	A	
F6	A	
F7	C	Arbiter was aware of whether report was by the reference standard or an observer under evaluation
F8	A	
F9	A	
F10	A	
F11	A	
G1	A	
H1	A	
H2	N/A	
H3	A	
H4	A1	Double or triple blind consultant radiologist report
H5	A2	
H6	N/A	
I1	N/A	
I2	A	
I3	A	

Appendix 8 Completed checklist - Study 6

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	**Recommendation**	**Complete /considere**
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	✓
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	✓
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	✓
Objectives	3	State specific objectives, including any prespecified hypotheses	✓ ?
Methods			
Study design	4	Present key elements of study design early in the paper	✓
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	✓ ?
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	✓ ?
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	✓ ?
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	✓
Bias	9	Describe any efforts to address potential sources of bias	
Study size	10	Explain how the study size was arrived at	x
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	x
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	✓
		(*b*) Describe any methods used to examine subgroups and interactions	✓
		(*c*) Explain how missing data were addressed	✓
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	✓
		(*e*) Describe any sensitivity analyses	

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	✓
		(b) Give reasons for non-participation at each stage	✓
		(c) Consider use of a flow diagram	x
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	x ?
		(b) Indicate number of participants with missing data for each variable of interest	✓
Outcome data	15*	Report numbers of outcome events or summary measures	x
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	✓
		(*b*) Report category boundaries when continuous variables were categorized	✓
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	x
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	✓
Discussion			
Key results	18	Summarise key results with reference to study objectives	✓
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	✓
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	✓
Generalisability	21	Discuss the generalisability (external validity) of the study results	✓
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	✓

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Appendix 9 Completed checklist – Study 7

Piper, K., Buscall, K., & Thomas, N. (2010). MRI reporting by radiographers: Findings of an accredited postgraduate programme. *Radiography*, *16*(2), 136-142

Study assessed using Diagnostic accuracy criteria (Appendix 3)

Criteria	Score	Comments
A1	A	
A2	B	
C1	A	
C2	A	
D1	A	
D2	N/A	
D3	A	
E1	N/A	
E2	A	
E3	N/A	
E4	C	Intra-observer variability not assessed
E5	A	
E6	B	Intra-arbiter variability not assessed
E7	C	Single arbiter – see E6
F1	A	
F2	A	
F3	A	
F4	A	
F5	A	
F6	A	
F7	C	Arbiter was aware of whether report was by the reference standard or an observer under evaluation
F8	A	
F9	A	
F10	A	
F11	A	
G1	A	
H1	A	
H2	N/A	
H3	A	
H4	A1	Double or triple blind consultant radiologist report
H5	A2	
H6	N/A	
I1	N/A	
I2	A	
I3	A	

Appendix 10 Completed checklist – Study 8

Brealey, S., K. Piper, D. King, M. Bland, J. Caddick, P. Campbell, A. Gibbon et al. (2013). Observer agreement in the reporting of knee and lumbar spine magnetic resonance (MR) imaging examinations: Selectively trained MR radiographers and consultant radiologists compared with an index radiologist. *European Journal of Radiology,* 82(10), e597-e605

Study assessed using Diagnostic outcome criteria (Appendix 3)

Criteria	Score	Comments
B1	A	
B2	A	
B3	A	
B4	A	
B5	A	
C1	A	
C2	A	
D1	A	
D2	A	
D3	A	
E1	N/A	
E2	A	
E3	N/A	
E4	C	Intra-observer variability not assessed
E5	A	
E6	B	Intra-arbiter variability not assessed
E7	C	Single arbiter – see E6
F1	A	Unaware of index standard report*
F2	A	Index standard unaware of observers reports*
F3	A	Could discuss with colleagues consistent with clinical practice*
F4	A	
F5	A	
F6	A	
F7	C	Arbiter compared observer reports to the index report*
F8	A	
F9	A	
F10	A	Access to relevant MRI study*
F11	A	
G1	A	
H1	A	
H2	N/A	
H3	A	
H4	A2	Expert experienced MSK consultant radiologist*
H5	A2	
H6	N/A	
I1	A	
I2	C	Diagnostic accuracy not assessed*
I3	A	